Cohort Sampling Schemes for the Mantel-Haenszel Estimator: Extensions to multilevel covariates, stratified models, and robust variance estimators * †‡

Larry Goldstein and Bryan Langholz
University of Southern California

February 2, 2008

Abstract

In many epidemiological contexts, disease occurrences and their rates are naturally modelled by counting processes and their intensities, allowing an analysis based on martingale methods. These methods lend themselves to extensions of nested case-control sampling designs where general methods of control selection can be easily incorporated. This same methodology allows for extensions of the Mantel-Haenszel estimator in two main directions. First, a variety of new sampling designs can be incorporated which can yield substantial efficiency gains over simple random sampling. Second, the extension allows for the treatment of multiple level time dependent exposures.

1 Introduction

Mantel-Haenszel estimators (Mantel and Haenszel (1959)) have long been used in medical research to quantify one group's risk of disease relative to

*AMS 2000 subject classifications. 62N02, 62D05, 60F05.

†Key words and phrases: survival analysis, counting process, counter matching.

‡This work was supported by United States National Cancer Institute grant CA14089.

another. An excellent review of the development of the Mantel-Haenszel estimator for analysis of epidemiologic case-control studies, as well as the prominent role it has played in epidemiologic research generally, is given in Breslow (1996).

In this paper we consider Mantel-Haenszel estimators for nested case-control studies in which controls are sampled from risk sets determined by the cohort failure times (see e.g., Langholz and Goldstein (1996)). In recent work, Zhang, Fujii, and Yanagawa (2000) defined generalized Mantel-Haenszel estimators when controls are a simple random sample from the risk set and derived the properties of the estimator for right censored cohort data. Further Zhang (2000) developed estimators for a number of methods of sampling controls including sampling with and without replacement and geometric sampling and showed their consistency. We expand on the work of these authors by providing estimators for the entire class of control sampling methods considered by Borgan, Goldstein, and Langholz (1995) , defining a natural "least squares" extension of the dichotomous covariate Mantel-Haenszel estimator to a multi-level covariate, and providing estimators of baseline hazard when a Mantel-Haenszel estimator is used for estimation of the rate ratio. Further, we show the consistency and asymptotic normality of these estimators under very general conditions, provide a number examples including random sampling, matching, and counter-matching and use the asymptotic variance results to compare the variance of the Mantel-Haenszel estimator to that of the maximum partial likelihood estimator, or MPLE. Moreover, at the end of Section 3, we show that our extension of the classical Mantel-Haenszel estimator in the dichotomous exposure situation has the same asymptotic variance at the null as the MPLE, for all such sampling schemes, in general.

In a cohort $\mathcal{R} = \{1, \ldots, n\}$ of individuals followed over a time interval $[0, \tau]$ with $0 < \tau \leq \infty$, a natural model relating failure and a binary exposure Z is that the failure rate for individuals $i \in \mathcal{R}$ with exposure covariate $Z_i = 1$ (group 1) is increased by an unknown factor $\phi_0 \in (0, \infty)$ over the failure rate for those unexposed, with covariate $Z_i = 0$ (group 0). The Mantel-Haenszel estimator for event time data provides a consistent and asymptotically normal estimate of the factor ϕ_0 in the semi-parametric model where individuals share a common but unknown baseline hazard function $\lambda_0(t)$ and fail at rate $\lambda_0(t)\phi_0^Z$ (Robins et al. (1986)). Letting $\mathcal{T}_j$ be the collection of all failure times among the individuals in group j, $n_k(t)$ the number of individuals in group k at time t and $n(t)$ the total number of

individuals at risk at time t, with

$$R_{jk} = \sum_{t \in \mathcal{T}_j} \frac{n_k(t)}{n(t)}, \tag{1}$$

the classical Mantel-Haenszel estimator is explicitly given by

$$\hat{\phi}_n = \frac{R_{10}}{R_{01}}. \tag{2}$$

It is well known that in the full cohort setting, the Mantel-Haenszel estimator (2) performs as well as the partial likelihood estimator at the null $\phi_0 = 1$. One contribution of this work is to show that this property is maintained when comparing these same two approaches under sampling, and provides our first reason to study the Mantel-Haenszel estimator. Secondly, we see by (2) that the classical Mantel-Haenszel estimator, computed from a cohort consisting of exposed and unexposed individuals, can be given "in closed form" without requiring the solution of a non-linear estimating equation, which must be done numerically. Again, this property of the estimator still prevails when sampling. A third reason to study the Mantel-Haenszel estimator is its popularity, which continues despite its efficiency drawbacks away from the null. For instance, a medline search of papers in the years 2000-2005 gives a total of 420 references where Mantel-Haenszel is cited in the abstract as the method applied. Since this methodology is quite popular, there is value in adapting it to sampling schemes like counter matching, which make the estimator much more efficient than its present version. Lastly, we cite Breslow (1996), himself quoting from page 156 of Kahn and Sempos (1989), "As Kahn and Sempos rightly remarked in their 1989 textbook Statistics in Epidemiology, when a method is as simple and free of assumptions as the Mantel-Haenszel procedure, it deserves a strong recommendation, and we do not hesitate to give it."

Although our main focus is on the use the Mantel-Haenszel estimator with various sampling schemes, we also extend its scope of applicability. In particular, suppose that to each individual $i \in \mathcal{R}$ there is assigned a time dependent covariate $Z_i(t)$ with values in $\{\alpha_0, \alpha_1, \ldots, \alpha_\eta\}$, and an indicator $Y_i(t)$ that equals one when i is observed, and zero otherwise. Letting the failure rate $\lambda_i(t)$ of individual i at time t equal

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\phi_0^{Z_i(t)}, \tag{3}$$

where $\lambda_0(t)$ is an unknown baseline hazard function, gives a model which accommodates multi-level exposure, censoring, and time dependent covariates. By incorporating a constant factor into $\lambda_0(t)$ if necessary, we may assume without loss of generality that $0 = \alpha_0 < \cdots < \alpha_\eta$. At any time t, the collection of individuals at risk

$$\mathcal{R}(t) = \{i : Y_i(t) = 1\}$$

may be divided into the $\eta + 1$ groups,

$$\mathcal{R}_k(t) = \{i \in \mathcal{R}(t) : Z_i(t) = \alpha_k\} \quad \text{with sizes} \quad n_k(t) = |\mathcal{R}_k(t)|, \quad k = 0, \ldots, \eta.$$

The individuals in $\mathcal{R}_k(t)$ for $k \neq 0$ are said to be exposed, and have an increased risk of $\phi_0^{\alpha_k}$ over those in $\mathcal{R}_0(t)$. The classical model under which the Mantel-Haenszel estimator has been developed is the case $\eta = 1, \alpha_1 = 1$.

In many practical situations, sampling schemes are necessary to accommodate situations where the collection of data in the full cohort $\mathcal{R}$ is impractical, expensive, or impossible. In general a cohort sampling scheme is given by specifying for all $i \in \mathbf{r} \subset \mathcal{R}$ a collection of probabilities $\pi_t(\mathbf{r}|i)$ for choosing the individuals in the set $\mathbf{r} \subset \mathcal{R}(t)$ to serve as controls should i fail at time t; we may set $\pi_t(\mathbf{r}|i) = 0$ when $i \notin \mathbf{r}$ or if i is not at risk at time t. The flexibility one can gain by the choice of design $\pi_t(\mathbf{r}|i)$ is substantial, opening up the possibility of using sampling designs that can take advantage of the structure of the data, resulting in substantial increases in efficiency.

Each design $\pi_t(\mathbf{r}|i)$ has an associated probability distribution on the subsets of $\mathcal{R}$ defined by

$$\pi_t(\mathbf{r}) = n(t)^{-1} \sum_{i \in \mathbf{r}} \pi_t(\mathbf{r}|i), \tag{4}$$

which sums to one by virtue of

$$\sum_{\mathbf{r} \subset \mathcal{R}} \sum_{i \in \mathbf{r}} \pi_t(\mathbf{r}|i) = \sum_{i \in \mathcal{R}} \sum_{\mathbf{r} \subset \mathcal{R}, \mathbf{r} \ni i} \pi_t(\mathbf{r}|i) = \sum_{i \in \mathcal{R}} Y_i(t) = n(t). \tag{5}$$

In addition, we can define the associated weights $w_i(t, \mathbf{r})$, set to 0 when i is not at risk, by

$$w_i(t, \mathbf{r}) = \frac{\pi_t(\mathbf{r}|i)}{n(t)^{-1} \sum_{l \in \mathbf{r}} \pi_t(\mathbf{r}|l)}, \quad \text{so that} \quad \pi_t(\mathbf{r}|i) = \pi_t(\mathbf{r}) w_i(t, \mathbf{r}). \tag{6}$$

We highlight a few sampling designs:

Design 1 *The Full Cohort. When information on all subjects is available, we may take* $\pi_t(\mathbf{r}|i)$ *to be the indicator of the set of those at risk at time* t*;*

$$\pi_t(\mathbf{r}|i) = \mathbf{1}(\mathbf{r} = \mathcal{R}(t)) \quad \textit{and so} \quad w_i(t, \mathbf{r}) = \mathbf{1}(i \in \mathcal{R}(t), \mathbf{r} = \mathcal{R}(t)).$$

The classical Mantel-Haenszel estimator is recovered under this scheme when $\eta = 1$ *and the covariates are time fixed. More generally, in this design and others, we allow censoring and multi-level, time dependent exposures.*

When the collection of covariate data on the full cohort is impractical and no additional information on cohort members is available, the nested case-control design is a natural choice:

Design 2 *Nested Case-Control Sampling. At each failure time, a simple random sample of* $m-1$ *individuals is chosen from those at risk to serve as controls for the failure;*

$$\pi_t(\mathbf{r}|i) = \binom{n(t)-1}{m-1}^{-1} \mathbf{1}(\mathbf{r} \subset \mathcal{R}(t), \mathbf{r} \ni i, |\mathbf{r}| = m).$$

The probabilities in (4) and weights (6) for this design are given, respectively, by

$$\pi_t(\mathbf{r}) = \binom{n(t)}{m}^{-1} I(\mathbf{r} \subset \mathcal{R}(t), |\mathbf{r}| = m) \quad \textit{and} \quad w_i(t, \mathbf{r}) = \frac{n(t)}{m},$$

for $i \in \mathbf{r} \subset \mathcal{R}(t)$*.*

The next two designs we consider, matching and counter matching, depend on the availability of some additional information on all cohort members. In particular, we assume that for each $i \in \mathcal{R}(t)$ we have available the value $C_i(t)$ giving the strata membership of i among the possible values in $\mathcal{C}$, some (small) finite set. For $l \in \mathcal{C}$ let

$$\mathcal{C}_l(t) = \{i : Y_i(t) = 1, C_i(t) = l\} \quad \text{and} \quad c_l(t) = |\mathcal{C}_l(t)|,$$

the l^{th} sampling stratum, and its size, at time t.

Design 3 *Matching, with specification* $\mathbf{m} = (m_l)_{l \in \mathcal{C}}$*,* $m_l \geq 1$*. If subject* i *fails at time* t*, then a simple random sample of* $m_{C_i(t)} - 1$ *controls are chosen*

from $\mathcal{C}_{C_i(t)}(t)$, the failure's stratum at time t, to serve as controls for the failure. Hence, the sampling probabilities of this scheme are given by

$$\pi_t(\mathbf{r}|i) = \binom{c_{C_i(t)}(t) - 1}{m_{C_i(t)} - 1}^{-1} \mathbf{1}(\mathbf{r} \subset \mathcal{C}_{C_i(t)}(t), \mathbf{r} \ni i, |\mathbf{r}| = m_{C_i(t)}).$$

The probabilities in (4) and weights (6) for this design are given, respectively, by

$$\pi_t(\mathbf{r}) = \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)} \binom{c_l(t)}{m_l}^{-1} I(\mathbf{r} \subset \mathcal{C}_l(t), |\mathbf{r}| = m_l)$$

and

$$w_i(t, \mathbf{r}) = n(t) \sum_{l \in \mathcal{C}} \frac{1}{m_l} I(\mathbf{r} \subset \mathcal{C}_l(t), |\mathbf{r}| = m_l, i \in \mathbf{r}).$$

The matching design could be used to control for confounding by stratifying by a potential confounder. In this design we can apply the estimator (1) with no change in the more general situation where there is a different baseline hazard in each strata. The consistency of $\hat{\phi}_n$ in this situation is preserved when the various conditions are satisfied in each separate strata. For details, and the asymptotic variance in this case, see the analysis of this design in Section 5.

Design 4 *Counter Matching, with specification $\mathbf{m} = (m_l)_{l \in \mathcal{C}}$, $m_l \geq 1$. If subject i fails at time t, then m_l controls are randomly sampled without replacement from each $\mathcal{C}_l(t)$ except for the failure's stratum, from which $m_{C_i(t)} - 1$ controls are sampled. Let $\mathcal{P}_{\mathcal{C}}(t)$ denote the set of all subsets of $\mathcal{R}(t)$ with m_l individuals of type l for all $l \in \mathcal{C}$. Then for $\mathbf{r} \in \mathcal{P}_{\mathcal{C}}(t)$ and $i \in \mathbf{r}$, the sampling probabilities of this scheme are given by*

$$\pi_t(\mathbf{r}|i) = \left[\prod_{l \in \mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1} \frac{c_{C_i(t)}(t)}{m_{C_i(t)}}.$$

The probabilities in (4) and weights (6) for this design are given, respectively, by

$$\pi_t(\mathbf{r}) = \left[\prod_{l \in \mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1} I(\mathbf{r} \subset \mathcal{R}(t), |\mathbf{r} \cap \mathcal{C}_l(t)| = m_l;\ l \in \mathcal{C})$$

and

$$w_i(t, \mathbf{r}) = c_{C_i(t)}(t)/m_{C_i(t)}.$$

An important instance where the counter matching design can be applied is where a surrogate exposure is available on all subjects. In Section 5, we show that significant efficiency gains over random sampling can be achieved when the surrogate exposure is sufficiently correlated with the true. Additional sampling schemes for which our results can be applied can be found in Borgan, Goldstein, and Langholz (1995), in particular, counter matching with additionally randomly sampled controls, and quota sampling.

To study sampling schemes and provide an extension of the Mantel-Haenszel estimator which functions in a generality that accommodates time varying and multi-level exposures, we set the model in the counting process framework. Let $N_{i,\mathbf{r}}(t)$ be the counting process that records the number of times in $(0,t]$ that i fails and $\mathbf{r}$ is chosen as the sampled risk set. By summing the counting processes $N_{i,\mathbf{r}}(t)$, we obtain

$$N_{\mathbf{r}}^k(t) = \sum_{i \in \mathcal{R}_k(t)} N_{i,\mathbf{r}}(t) \quad \text{and} \quad N_{\mathbf{r}}(t) = \sum_{i \in \mathbf{r}} N_{i,\mathbf{r}}(t), \tag{7}$$

recording, respectively, the number of times in $(0,t]$ that $\mathbf{r}$ was chosen as the sampled risk set for a failure in $\mathcal{R}_k(t)$, and the total number of times in $(0,t]$ that $\mathbf{r}$ was chosen as the sampled risk. Now let

$$A_{\mathbf{r}}^k(t) = \sum_{i \in \mathcal{R}_k(t)} \pi_t(\mathbf{r}|i) = \pi_t(\mathbf{r}) \sum_{i \in \mathcal{R}_k(t)} w_i(t,\mathbf{r}), \quad k = 0, \ldots, \eta. \tag{8}$$

For a given continuous function $a : \mathbf{R}^{\eta+1} \to [0,\infty)$, define

$$a_{\mathbf{r}}(t) = a(A_{\mathbf{r}}^0(t), \ldots, A_{\mathbf{r}}^\eta(t)),$$

and suppressing dependence on a, for $j \neq k$ set

$$R_{jk}(t) = \int_0^t \sum_{\mathbf{r} \subset \mathcal{R}} a_{\mathbf{r}}(s) A_{\mathbf{r}}^k(s) dN_{\mathbf{r}}^j(s), \quad \text{and} \quad R_{jk} = R_{jk}(\tau). \tag{9}$$

It is convenient to choose an a for which

$$|a(v_0, \ldots, v_\eta) v_k| \leq 1 \quad \text{for all } v_k \geq 0, k = 0, \ldots, \eta. \tag{10}$$

A natural choice which satisfies condition (10) is

$$a(v_0, \ldots, v_\eta) = (v_0 + \cdots + v_\eta)^{-1}, \tag{11}$$

extending the $\eta = 1$ canonical choice of $a(u, v) = (u + v)^{-1}$. By (4), for this a and sets $\mathbf{r}$ with $\pi_t(\mathbf{r}) \neq 0$ we have

$$a_{\mathbf{r}}(t) = \left(\sum_{i \in \mathcal{R}} \pi_t(\mathbf{r}|i) \right)^{-1} = (n(t)\pi_t(\mathbf{r}))^{-1}, \tag{12}$$

and hence, by (9), with $t_{1,j} < t_{2,j} \ldots$ the ordered collection of failure times for individuals with covariate j at the time of failure, and $\widetilde{\mathcal{R}}_{l,j}$ the sampled risk set at failure time $t_{l,j}$,

$$R_{jk} = \sum_{l \geq 1} \frac{1}{n(t)} \sum_{i \in \widetilde{\mathcal{R}}_{l,j}, Z_i(t_{l,j}) = k} w_i(t, \widetilde{\mathcal{R}}_{l,j}). \tag{13}$$

For the full cohort information (Design 1), since

$$\widetilde{\mathcal{R}}_{l,j} = \mathcal{R}(t_{l,j}) \quad \text{and} \quad w_i(t, \mathcal{R}(t)) = 1 \quad \text{for } i \in \mathcal{R}(t),$$

the expression for R_{jk} in (13) reduces to that in (1).

Noting that for $\eta = 1$ the estimator (2) is the solution of the linear estimating equation

$$\phi R_{01} - R_{10} = 0,$$

it is therefore the unique minimizer of $G_{01}^2(\phi)$ where with $j < k$ we let

$$G_{jk}(\phi) = \phi^{\alpha_k} R_{jk} - \phi^{\alpha_j} R_{kj}. \tag{14}$$

Hence, given non-negative constants c_{jk} not all zero, we propose as our estimator a value $\hat{\phi}_n$ which minimizes the weighted sum of squares

$$n^{-1} \sum_{j<k} c_{jk} G_{jk}^2(\phi),$$

that is, a solution to the estimating equation $\mathcal{U}_n(\phi) = 0$, where, with $G'_{jk}(\phi)$ denoting the derivative of $G_{jk}(\phi)$ with respect to ϕ,

$$\begin{aligned} \mathcal{U}_n(\phi) &= n^{-1} \sum_{j<k} c_{jk} G_{jk}(\phi) G'_{jk}(\phi) \qquad (15) \\ &= n^{-1} \sum_{j<k} c_{jk} (\phi^{\alpha_k} R_{jk} - \phi^{\alpha_j} R_{kj})(\alpha_k \phi^{\alpha_k - 1} R_{jk} - \alpha_j \phi^{\alpha_j - 1} R_{kj}). \end{aligned}$$

We prove that the estimator $\hat{\phi}_n$ is consistent for ϕ_0 under the conditions specified in Theorem 3.1, and establish its asymptotic normal distribution in Theorem 3.2. Proposition 3.3 shows how to choose the constants c_{jk} to achieve the minimum asymptotic variance over the class of all estimators of this form. For other possibilities regarding the construction of estimating equations which may have some efficiency advantages, see Qu et al. (2000), Godambe (1960), and Heyde (1997).

Where estimates of ϕ_0 can be used to assess the magnitude of the effect that exposure has on failure, estimates of the integrated baseline hazard

$$\Lambda_0(t) = \int_0^t \lambda_0(u)du$$

can in turn be used to provide estimates of absolute risk. We consider the integrated baseline hazard function estimate

$$\hat{\Lambda}_n(t, \hat{\phi}_n) = \int_0^t \sum_{\mathbf{r} \subset \mathcal{R}} \frac{dN_r(u)}{\sum_{i \in \mathbf{r}} \hat{\phi}_n^{Z_i(t)} w_i(u, \mathbf{r})}, \tag{16}$$

given in terms of the weights defined in (6), where the ratio in the integral is regarded as 0 if there is no one at risk. In Theorem 4.1 we give conditions under which

$$\sqrt{n}\left(\hat{\Lambda}_n(\cdot, \hat{\phi}_n) - \Lambda(\cdot)\right)$$

converges weakly as $n \to \infty$ to a mean zero Gaussian process, and provide a uniformly consistent estimator for its variance function.

The counting process model and some of its consequences are derived in Section 2. The consistency and asymptotic normality of $\hat{\phi}_n$ and $\hat{\Lambda}_n$ are proved in Sections 3 and 4 respectively. In Section 5 we study the asymptotic properties of these estimators under Designs 1 - 4, and present efficiency comparisons against the partial likelihood estimator. Much of the analysis here follows the work of Borgan, Goldstein, and Langholz (1995) closely, and is hereafter referred to as BGL.

2 The Counting Process Model for Sampling

We will assume that the censoring and failure information are defined on a probability space with a standard filtration $\mathcal{F}_t$, and that the censoring

indicators $Y_i(t)$, exposures $Z_i(t)$, design $\pi_t(\mathbf{r}|i)$ and strata variables $C_i(t)$ are left continuous and adapted, and hence predictable and locally bounded. We make the assumption of independent sampling as in BGL, that the intensity processes with respect to the filtration $\mathcal{F}_t$ is the same as that with respect to this filtration augmented with the sampling information; in other words, we assume that selecting an individual as a control does not influence the likelihood of failure of the individual in the future. We assume that the intensity process of $N_{i,\mathbf{r}}(t)$ is given by

$$\lambda_{i,\mathbf{r}}(t) = \phi_0^{Z_i(t)} \pi_t(\mathbf{r}|i)\lambda_0(t), \tag{17}$$

so that subtracting the integrated intensity from the counting processes $N_{i,\mathbf{r}}(t)$ results in the orthogonal local square integrable martingales

$$M_{i,\mathbf{r}}(t) = N_{i,\mathbf{r}}(t) - \int_0^t \lambda_{i,\mathbf{r}}(s)ds, \tag{18}$$

with predictable quadratic variation

$$d < M_{i,\mathbf{r}} >_t = \lambda_{i,\mathbf{r}}(t)dt.$$

Further, we assume that the baseline hazard function $\lambda_0(t)$ is bounded away from zero and infinity.

With $A_{\mathbf{r}}^k(t)$ given in (8), by linearity the counting processes $N_{\mathbf{r}}^k(t)$ and $N_{\mathbf{r}}(t)$ defined in (7) have respective intensities

$$\lambda_{\mathbf{r}}^k(t) = \phi_0^{\alpha_k} A_{\mathbf{r}}^k(t)\lambda_0(t) \quad \text{and} \quad \lambda_{\mathbf{r}}(t) = \sum_{k=0}^{\eta} \phi_0^{\alpha_k} A_{\mathbf{r}}^k(t)\lambda_0(t), \tag{19}$$

and give rise to the orthogonal local square integrable martingales

$$\begin{aligned} M_{\mathbf{r}}^k(t) &= \sum_{i\in\mathcal{R}_k(t)} M_{i,\mathbf{r}}(t) = N_{\mathbf{r}}^k(t) - \int_0^t \lambda_{\mathbf{r}}^k(s)ds \quad \text{and} \\ M_{\mathbf{r}}(t) &= \sum_{i\in\mathbf{r}} M_{i,\mathbf{r}}(t) = N_{\mathbf{r}}(t) - \int_0^t \lambda_{\mathbf{r}}(s)ds, \end{aligned}$$

with predictable variations

$$d < M_{\mathbf{r}}^k >_t = \lambda_{\mathbf{r}}^k(t)dt \quad \text{and} \quad d < M_{\mathbf{r}} >_t = \lambda_{\mathbf{r}}(t)dt.$$

Using (9), (18), (8) and (19), we have

$$R_{jk}(t) = \int_0^t \sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{r}}(s)A^k_{\mathbf{r}}(s)\left(\phi_0^{\alpha_j}A^j_{\mathbf{r}}(s)\lambda_0(s)ds + dM^j_{\mathbf{r}}(s)\right)$$

$$= \phi_0^{\alpha_j}\int_0^t \sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{r}}(s)A^k_{\mathbf{r}}(s)\ A^j_{\mathbf{r}}(s)\ \lambda_0(s)ds + \int_0^t \sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{r}}(s)A^k_{\mathbf{r}}(s)dM^j_{\mathbf{r}}(s) \quad (20)$$

For $\mathbf{v}$ a multi-subset of $\{0,\ldots,\eta\}$, e.g. $\mathbf{v}=\{0,0,1\}$, let

$$H_{\mathbf{v}}(t) = \sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t)\prod_{k\in\mathbf{v}} A^k_{\mathbf{r}}(t). \quad (21)$$

In particular, for $|\mathbf{v}|=2$, and subscripting by jk rather than $\{j,k\}$ for notational convenience, we have

$$H_{jk}(t) = \sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{r}}(t)A^j_{\mathbf{r}}(t)A^k_{\mathbf{r}}(t).$$

Letting in addition

$$W_{jk}(t) = \int_0^t \sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{r}}(s)A^k_{\mathbf{r}}(s)dM^j_{\mathbf{r}}(s), \quad (22)$$

we may write (20) as

$$R_{jk}(t) \quad = \quad \phi_0^{\alpha_j}\int_0^t H_{jk}(s)\lambda_0(s)ds + W_{jk}(t). \quad (23)$$

The processes W_{jk} are local square integrable martingales, and by the orthogonality of $M^j_{\mathbf{r}}(s)$ and (19), have predictable quadratic covariation

$$\begin{aligned} d<W_{jk},W_{pq}>_t \quad &= \quad \mathbf{1}_{(j=p)}\sum_{\mathbf{r}\subset\mathcal{R}} a^2_{\mathbf{r}}(t)A^k_r(t)A^q_r(t)\lambda^j_{\mathbf{r}}(t)dt \\ &= \quad \mathbf{1}_{(j=p)}\phi_0^{\alpha_j}H_{jkq}(t)\lambda_0(t)dt, \end{aligned} \quad (24)$$

so in particular

$$d<W_{jk}>_t = \phi_0^{\alpha_j}H_{jkk}(t)\lambda_0(t)dt. \quad (25)$$

By (23) and $H_{jk} = H_{kj}$ we have, suppressing the dependence on ϕ which is explicit in (14), now considering G_{jk} a function of t, for $j < k$ we have

$$G_{jk}(t) = \phi_0^{\alpha_k} R_{jk}(t) - \phi_0^{\alpha_j} R_{kj}(t) = \phi_0^{\alpha_k} W_{jk}(t) - \phi_0^{\alpha_j} W_{kj}(t) \tag{26}$$

are local square integrable martingales with quadratic covariation, for $j < k$, $p < q$, by (24),

$$\begin{aligned}
& d < G_{jk}, G_{pq} >_t = d < \phi_0^{\alpha_k} W_{jk} - \phi_0^{\alpha_j} W_{kj}, \phi_0^{\alpha_q} W_{pq} - \phi_0^{\alpha_p} W_{qp} >_t \\
= & \Big(\mathbf{1}_{(j=p)} \phi_0^{\alpha_k+\alpha_q+\alpha_j} H_{jkq}(t) - \mathbf{1}_{(j=q)} \phi_0^{\alpha_k+\alpha_p+\alpha_j} H_{jkp}(t) \\
& - \mathbf{1}_{(k=p)} \phi_0^{\alpha_j+\alpha_q+\alpha_k} H_{kjq}(t) + \mathbf{1}_{(k=q)} \phi_0^{\alpha_j+\alpha_p+\alpha_k} H_{kjp}(t) \Big) \lambda_0(t) dt \\
= & \phi_0^{\alpha_j+\alpha_k} \left(\left(\mathbf{1}_{(j=p)} - \mathbf{1}_{(k=p)} \right) \phi_0^{\alpha_q} H_{jkq}(t) + \left(\mathbf{1}_{(k=q)} - \mathbf{1}_{(j=q)} \right) \phi_0^{\alpha_p} \right) H_{jkp}(t) \lambda_0(t) dt;
\end{aligned}$$

in particular,

$$d < G_{jk} >_t = \phi_0^{\alpha_k+\alpha_j} \left(\phi_0^{\alpha_k} H_{jkk}(t) + \phi_0^{\alpha_j} H_{kjj}(t) \right) \lambda_0(t) dt.$$

3 Asymptotics of $\hat{\phi}_n$

We prove the consistency and asymptotic normality of $\hat{\phi}_n$, a solution to $\mathcal{U}_n(\phi) = 0$ with $\mathcal{U}_n(\phi)$ given by (15), under some regularity and stability conditions.

Condition 1 *The cumulative hazard on the interval* $[0, \tau]$ *is finite:*

$$\Lambda_0(\tau) < \infty.$$

For $H_{\mathbf{v}}(t)$ given in (21), define

$$h_{n,\mathbf{v}}(t) = \frac{1}{n} H_{\mathbf{v}}(t). \tag{27}$$

Condition 2 *For* $h_{n,\mathbf{v}}(t)$ *with* $|\mathbf{v}| \in \{2, 3\}$*, there exist left continuous functions* $\bar{h}_{n,\mathbf{v}}(t), h_{\mathbf{v}}(t), \bar{h}_{\mathbf{v}}(t)$ *such that for all* $t \in [0, \tau]$*,*

$$0 \le h_{n,\mathbf{v}}(t) \le \bar{h}_{n,\mathbf{v}}(t),$$

and for almost all t in $[0,\tau]$,

$$h_{n,\mathbf{v}}(t) \to_p h_{\mathbf{v}}(t), \quad \text{and} \quad \bar{h}_{n,\mathbf{v}}(t) \to_p \bar{h}_{\mathbf{v}}(t),$$

and

$$\int_0^\tau \bar{h}_{n,\mathbf{v}}(t)\lambda_0(t)dt \to_p \int_0^\tau \bar{h}_{\mathbf{v}}(t)\lambda_0(t)dt < \infty.$$

Note that for any a satisfying (10), so in particular for the canonical choice of a given by (11), we may take $\bar{h}_{n,\mathbf{v}}(t) = \bar{h}_{\mathbf{v}}(t) = 1$ since for the $|\mathbf{v}|-1$ terms in the product in (21) we have $a_{\mathbf{r}}(t)A_{\mathbf{r}}^k(t) \leq 1$, and applying the additional factor $1/n$ granted in (27), we have by (5) that $A_{\mathbf{r}}^k(t)/n \leq 1$, taking care of the remaining factor of $A_{\mathbf{r}}^k(t)$ in the product. Hence, if Condition 1 holds and a satisfies (10) then Condition 2 holds provided only that $h_{n,\mathbf{v}}(t) \to_p h_{\mathbf{v}}(t)$ for $|\mathbf{v}| \in \{2,3\}$.

The following version of the dominated convergence theorem is due to Hjort and Pollard (1993) :

Proposition 3.1 *Suppose $\Lambda_0(\tau) < \infty$ and let $0 \leq U_n(t) \leq \bar{U}_n(t)$ be left-continuous random processes on the interval $[0,\tau]$. Suppose $\bar{U}_n(t) \to_p \bar{U}(t)$ and $U_n(t) \to_p U(t)$ for almost all t, as $n \to \infty$, and that $\int_0^\tau \bar{U}_n(s)\lambda_0(s)ds \to_p \int_0^\tau \bar{U}(s)\lambda_0(s)ds < \infty$. Then $\int_0^t U_n(s)\lambda_0(s)ds \to_p \int_0^t U(s)\lambda_0(s)ds$ for all $t \in [0,\tau]$ as $n \to \infty$.*

For given $\mathbf{v}$ and and corresponding $h_{\mathbf{v}}(t)$, let

$$I_{\mathbf{v}}(t) = \int_0^t h_{\mathbf{v}}(s)\lambda_0(s)ds, \tag{28}$$

and for p a non-negative integer, let $(\alpha)_p = \alpha!/(\alpha-p)!$, the falling factorial.

Proposition 3.2 *Let Conditions 1 and 2 hold. Then for every $t \in [0,\tau]$,*

$$n^{-1} < W_{jk}, W_{pq} >_t \to_p \mathbf{1}_{(j=p)}\phi_0^{\alpha_j} I_{jkq}(t), \tag{29}$$

and

$$n^{-1}R_{jk}(t) \to_p \phi_0^{\alpha_j} I_{jk}(t). \tag{30}$$

Furthermore, as $n \to \infty$, for all $j < k$ and $p = 0,1,\ldots$, the p^{th} derivatives of $G_{jk}(\phi)$ defined in (14) satisfy

$$n^{-1}G_{jk}^{(p)}(\phi) \to_p ((\alpha_k)_p\phi^{\alpha_k-p}\phi_0^{\alpha_j} - (\alpha_j)_p\phi^{\alpha_j-p}\phi_0^{\alpha_k})I_{jk}(\tau). \tag{31}$$

In particular, for $p = 1$

$$n^{-1}G'_{jk}(\phi_0) \to_p \beta_{jk} \equiv (\alpha_k - \alpha_j)\phi_0^{\alpha_k+\alpha_j-1} I_{jk}(\tau), \tag{32}$$

and with $\mathcal{U}_n(\phi)$ as in (15),

$$n^{-1}\mathcal{U}'_n(\phi_0) \to_p \gamma \quad \textit{where} \quad \gamma = \sum_{j<k} c_{jk}\beta_{jk}^2. \tag{33}$$

Proof: Conditions 1 and 2 and Proposition 3.1 give

$$\int_0^t h_{n,\mathbf{v}}(s)\lambda_0(s)ds \to_p \int_0^t h_{\mathbf{v}}(s)\lambda_0(s)ds \tag{34}$$

for $|\mathbf{v}| \in \{2, 3\}$. In particular, with $|\mathbf{v}| = 3$, by (24) we obtain (29).

By (23) and (27) we have

$$\frac{1}{n}R_{jk}(t) = \phi_0^{\alpha_j} \int_0^t h_{n,jk}(s)\lambda_0(s)ds + \frac{1}{n}W_{jk}(t).$$

By (34) the first term converges to the right hand side of (29). For the second term, by (25), (27) and Lenglart's inequality (see Andersen et al. (1993)), for all positive ϵ, δ,

$$P\left(\sup_{t\le\tau} |\frac{1}{n}W_{jk}(t)| > \epsilon\right) \le \frac{\delta}{\epsilon^2} + P\left(\frac{\phi_0^{\alpha_j}}{n}\int_0^\tau h_{n,jkk}(t)\lambda_0(t)dt > \delta\right).$$

Now applying (34) for $\mathbf{v} = \{j, k, k\}$, we see

$$\frac{1}{n}W_{jk}(t) \to_p 0 \quad \text{as } n \to \infty,$$

and hence (30); (31) now follows immediately from (14) and (30).

Taking derivatives in (15) yields

$$n^{-1}\mathcal{U}'_n(\phi) = n^{-2}\sum_{j<k} c_{jk}\left((G'_{jk}(\phi))^2 + G_{jk}(\phi)G''_{jk}(\phi)\right),$$

and (33) now follows using (32) for the first term, and (31) for $p = 0$ at $\phi = \phi_0$ to show the second term varnishes. ■

Our first result gives the consistency of $\hat{\phi}_n$ under the following additional non-triviality condition; in particular, note that c_{jk} can always be chosen to be positive for all $j < k$.

Condition 3 *There exists some pair* $j < k$ *for which both* c_{jk} *in (15) and* $I_{jk}(\tau)$ *in (28) are strictly positive.*

Theorem 3.1 *Under Conditions 1 through 3, the estimating equation (15) has a consistent sequence of solutions*

$$\hat{\phi}_n \to_p \phi_0 \quad \textit{as } n \to \infty.$$

Proof: By the arguments of Aitchison and Silvey (1958) and Billingsley (1961), it suffices to show that as $n \to \infty$,

$$\begin{aligned} &n^{-1}\mathcal{U}_n(\phi_0) \to_p 0, && (35)\\ &n^{-1}\mathcal{U}_n'(\phi_0) \quad \text{converges in probability to a positive number,} \end{aligned}$$

and that there is a neighborhood Θ_0 of ϕ_0 such that for every $\eta \in (0,1)$, there is a K such that for all n,

$$P(\frac{1}{n}|\mathcal{U}_n''(\phi)| \leq K, \phi \in \Theta_0) \;\; \geq \;\; 1 - \eta. \tag{36}$$

Applying (31) for $p = 0, 1$ gives the first part of (35), the second part is (33), the positivity of γ following from Condition 3.

By (31), each term in the second derivative

$$n^{-1}\mathcal{U}''(\phi) = n^{-2}\sum_{j<k} c_{jk}(3G_{jk}'(\phi)G_{jk}''(\phi) + G_{jk}(\phi)G_{jk}'''(\phi))$$

is uniformly bounded in probability in any bounded neighborhood Θ_0 of ϕ_0 not containing zero, giving the uniform boundedness in probability condition (36). ■

To obtain the limiting distribution of $\hat{\phi}_n$ and $\hat{\Lambda}_n(\cdot)$ we assume the following

Condition 4 *There exists* $\delta > 0$ *such that for all* $j < k$,

$$\frac{1}{n^{1+\delta/2}}\int_0^\tau \sum_{\mathbf{r}\subset\mathcal{R}} |a_{\mathbf{r}}(t)A_{\mathbf{r}}^k(t)|^{2+\delta}A_{\mathbf{r}}^j(t)\lambda_0(t)dt \to_p 0.$$

Note that Condition 4 is satisfied for any $\delta > 0$ using any function a satisfying (10), so in particular the canonical function a given in (11), since by (5),

$$\sum_{\mathbf{r}\subset\mathcal{R}} |a_{\mathbf{r}}(t)A_{\mathbf{r}}^k(t)|^{2+\delta}A_{\mathbf{r}}^j(t) \leq \sum_{\mathbf{r}\subset\mathcal{R}} A_{\mathbf{r}}^j(t) \leq \sum_{\mathbf{r}\subset\mathcal{R}}\sum_{i\in\mathbf{r}} \pi_t(\mathbf{r}|i) = n(t) \leq n.$$

Lemma 3.1 *Under Conditions 1-4, the processes* $\{n^{-1/2}W_{jk}(\cdot)\}_{j,k}$ *given in (22) converge jointly in* $D[0,\tau]$ *to the mean zero Gaussian processes* $\{w_{jk}(\cdot)\}_{j,k}$ *with covariation function*

$$< w_{jk}, w_{pq} >_t = \mathbf{1}_{(j=p)} \phi_0^{\alpha_j} I_{jkq}(t),$$

and hence the collection $\{n^{-1/2}G_{jk}\}_{j<k}$ *given in (26) converges jointly in* $D[0,\tau]$ *to the mean zero Gaussian processes* $\{g_{jk}(\cdot)\}_{j<k}$ *with covariation function*

$$\begin{aligned} < & \ g_{jk}, g_{pq} >_t \\ = & \ \phi_0^{\alpha_j+\alpha_k+\alpha_q} \left(\mathbf{1}_{(j=p)} - \mathbf{1}_{(k=p)}\right) I_{jkq}(t) + \phi_0^{\alpha_j+\alpha_k+\alpha_p} \left(\mathbf{1}_{(k=q)} - \mathbf{1}_{(j=q)}\right) I_{jkp}(t), \end{aligned} \tag{37}$$

so in particular,

$$< g_{jk} >_t = \phi_0^{\alpha_k+\alpha_j} \int_0^t \left(\phi_0^{\alpha_k} h_{jkk}(s) + \phi_0^{\alpha_j} h_{kjj}(s)\right) \lambda_0(s) ds.$$

Further, for any permutation (ι, κ, χ) *of* (j,k,q)*, all* $t \in [0,\tau]$*, and any consistent sequence* $\hat{\phi}_n \to_p \phi_0$*, as* $n \to \infty$*,*

$$n^{-1}\hat{\phi}_n^{-\alpha_\iota}[W_{\iota\kappa}, W_{\iota\chi}]_t \to_p < w_{\iota\kappa}, w_{\iota\chi} >_t = I_{jkq}(t),$$

where

$$n^{-1}[W_{\iota\kappa}, W_{\iota\chi}]_t = \frac{1}{n}\sum_{\mathbf{r}\subset\mathcal{R}} \int_0^t a_{\mathbf{r}}^2(s) A_{\mathbf{r}}^{\kappa}(s) A_{\mathbf{r}}^{\chi}(s)\ dN_{\mathbf{r}}^{\iota}(s), \tag{38}$$

the scaled optional variation, so that

$$\hat{I}_{jkq}(t) \ = \ n^{-1}\sum \xi_{\iota\kappa\chi}\hat{\phi}_n^{-\alpha_\iota}[W_{\iota\kappa}, W_{\iota\chi}]_t \to_p I_{jkq}(t), \tag{39}$$

where the sum is over all permutations (ι,κ,χ) *of* (j,k,q)*, and* $\sum \xi_{\iota\kappa\chi} = 1$.

Proof: We apply the martingale central limit Theorem of Rebolledo, as presented in Theorem II.5.1 of Andersen et al. (1993). The processes $\{n^{-1/2}W_{jk}\}_{j,k}$ are local square integrable martingales, whose predictable quadratic variation converges by Proposition 3.2 to the continuous functions given in (29). Using the Lindeberg condition, Condition 4,

$$\begin{aligned} & \frac{1}{n}\int_0^\tau \sum_{\mathbf{r}\subset\mathcal{R}} \left(a_{\mathbf{r}}(t)A_{\mathbf{r}}^k(t)\right)^2 \mathbf{1}(n^{-1/2}|a_{\mathbf{r}}(t)A_{\mathbf{r}}^k(t)| > \epsilon)\lambda_{\mathbf{r}}^j(t)dt \\ \leq & \ \frac{\phi_0^{\alpha_j}}{\epsilon^\delta n^{1+\delta/2}} \int_0^\tau \sum_{\mathbf{r}\subset\mathcal{R}} |a_{\mathbf{r}}(t)A_{\mathbf{r}}^k(t)|^{2+\delta} A_{\mathbf{r}}^j(t)\lambda_0(t)dt \to_p 0. \end{aligned}$$

The convergence of the scaled optional variation (38) to the limit (29) of the scaled predictable variation follows from Theorem II.5.1 of Andersen et al. (1993). ■

With $S_{\mathbf{r}}^{(0)}(\phi, t)$ as in (52), in place of (38) one may consider the estimated scaled predictable variation

$$\frac{1}{n}\sum_{\mathbf{r}\subset\mathcal{R}}\int_0^t a_{\mathbf{r}}^2(s)A_{\mathbf{r}}^{\kappa}(s)A_{\mathbf{r}}^{\chi}(s)A_{\mathbf{r}}^{\iota}(s)\frac{dN_{\mathbf{r}}(s)}{S_{\mathbf{r}}^{(0)}(\hat{\phi}_n, s)}. \tag{40}$$

Since $dN_{\mathbf{r}}(t) = S_{\mathbf{r}}^{(0)}(\phi_0, t)\lambda_0(t) + dM_{\mathbf{r}}(t)$, replacing $\hat{\phi}_n$ by ϕ_0 in (40) gives the scaled predictable variation plus a martingale term, and so (40) converges in probability to $I_{jkq}(t)$ when the martingale term tends to zero and the replacement of $\hat{\phi}_n$ by ϕ_0 is asymptotically negligible.

The variance estimators based on (38) and (40) simplify considerably in special cases. The variance estimator of Zhang, Fujii, and Yanagawa (2000) for simple random sampling uses the estimated predictable variation (40). The empirical and conditional variance estimators of Breslow (1981), with one case per set, correspond to the optional and estimated predictable variation estimators for simple random sampling for the canonical a as in (11).

Theorem 3.2 *Under Conditions 1-4, for $\hat{\phi}_n$ any consistent sequence of solutions of the estimating equation (15), we have*

$$\sqrt{n}\left(\hat{\phi}_n - \phi_0\right) \to_d \mathcal{N}(0, \sigma^2) \quad \textit{where} \quad \sigma^2 = v^2/\gamma^2 \tag{41}$$

with γ as in (33) and

$$v^2 = \sum_{j<k,p<q} c_{jk}\beta_{jk} < g_{jk}, g_{pq} >_\tau \beta_{pq}c_{pq}, \tag{42}$$

with β_{jk} as in (32) and $< g_{jk}, g_{pq} >_t$ in (37). By (39) of Proposition 3.2, $< g_{jk}, g_{pq} >_\tau$ can be consistently estimated by $< \widehat{g_{jk}, g_{pq}} >_\tau$, given by

$$\hat{\phi}_n^{\alpha_j+\alpha_k+\alpha_q}\left(\mathbf{1}_{(j=p)} - \mathbf{1}_{(k=p)}\right)\hat{I}_{jkq}(\tau) + \hat{\phi}_n^{\alpha_j+\alpha_k+\alpha_p}\left(\mathbf{1}_{(k=q)} - \mathbf{1}_{(j=q)}\right)\hat{I}_{jkp}(\tau), \tag{43}$$

whereas by (32) and (30), β_{jk} can be consistently estimated by

$$\hat{\beta}_{jk} = (\alpha_k - \alpha_j)\hat{\phi}_n^{\alpha_k+\alpha_j-1}\hat{I}_{jk}(\tau) \tag{44}$$

where
$$\hat{I}_{jk}(t) = n^{-1} \sum_{\{j,k\}=\{\iota,\kappa\}} \xi_{\iota\kappa} \hat{\phi}_n^{-\alpha_\iota} R_{\iota\kappa}(t),$$
for any weights $\xi_{\iota\kappa}$ summing to one. Hence
$$\widehat{\sigma^2} = \frac{\widehat{v^2}}{\hat{\gamma}^2} \tag{45}$$
consistently estimates σ^2 where $\widehat{v^2}$ and $\hat{\gamma}$ are obtained by substituting (43) and (44) into (42), and (44) into (33) respectively.

In the parameterization $\theta = \log \phi$,
$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to \mathcal{N}(0, \sigma^2/\phi_0^2). \tag{46}$$

Proof: By the consistency of the solution $\hat{\phi}_n$ for ϕ_0, that $n^{-1}\mathcal{U}_n(\phi_0) \to_p \gamma > 0$, and the uniform boundedness of the second derivative of $\mathcal{U}_n(\phi)$ in a neighborhood of ϕ_0 given in (36), we have
$$\sqrt{n}(\hat{\phi}_n - \phi_0) = -\gamma^{-1} n^{-1/2} \mathcal{U}_n(\phi_0) + o_p(1). \tag{47}$$
But
$$n^{-1/2}\mathcal{U}_n(\phi_0) \quad \to_d \quad \mathcal{N}(0, v^2)$$
by (15), Lemma 3.1, and (32) of Proposition 3.2; (46) now follows by a direct application of the delta method. ■

Let **c** be the vector of the constants c_{jk} obtained by taking the pairs $j < k$ in some canonical order, say lexicographically; with the same indexing form a matrix Γ with entries $< g_{jk}, g_{pq} >_\tau$ and a matrix B with diagonal entries β_{jk}. Note that when Γ is positive definite the matrix $B\Gamma B$ is also, and therefore there exists a non-singular matrix M such that
$$B\Gamma B = M'M. \tag{48}$$
Proposition 3.3 provides the constants c_{jk} which minimize the asymptotic variance (41) of $\hat{\phi}_n$.

Proposition 3.3 *Let Γ be positive definite, **1** the vector all of whose entries are 1, M as in (48), and*
$$X = (M^{-1})' B^2 \mathbf{1}\mathbf{1}' B^2 M^{-1}.$$

Then taking

$$\mathbf{c} = M^{-1}\mathbf{d},$$

where $\mathbf{d}$ *is any eigenvector corresponding to the largest eigenvalue* λ *of* X*, minimizes the asymptotic variance (41) with the value* $\sigma^2 = \lambda^{-1}$*.*

Proof: In the given notation, we may write

$$\gamma = \mathbf{1}'B^2\mathbf{c} \quad \text{so that} \quad v^2 = \frac{\mathbf{c}'B\Gamma B\mathbf{c}}{\mathbf{c}'B^2\mathbf{1}\mathbf{1}'B^2\mathbf{c}}.$$

Then letting $\mathbf{d} = M\mathbf{c}$ we have by (48)

$$v^{-2} = \frac{\mathbf{c}'B^2\mathbf{1}\mathbf{1}'B^2\mathbf{c}}{\mathbf{c}'B\Gamma B\mathbf{c}} = \frac{\mathbf{c}'B^2\mathbf{1}\mathbf{1}'B^2\mathbf{c}}{\mathbf{c}'M'M\mathbf{c}} = \frac{\mathbf{d}'(M^{-1})'B^2\mathbf{1}\mathbf{1}'B^2M^{-1}\mathbf{d}}{\mathbf{d}'\mathbf{d}} = \frac{\mathbf{d}'X\mathbf{d}}{\mathbf{d}'\mathbf{d}},$$

which has its maximum value of λ, the largest eigenvalue of X, when $\mathbf{d}$ is a corresponding eigenvector. ■

For $\eta = 1$ and $\alpha_0 = 0, \alpha_1 = 1$, because the estimator $\hat{\phi}_n$ is given explicitly, the consistency and asymptotic normality of $\hat{\phi}_n$ can be shown in a more direct way, the framework, however, remains sufficiently general to include sampling. In particular, from (2) and Proposition 3.2,

$$\hat{\phi}_n = \frac{R_{10}}{R_{01}} = \frac{n^{-1}R_{10}}{n^{-1}R_{01}} \to_p \phi_0, \quad \text{as } n \to \infty,$$

and

$$\sqrt{n}\left(\hat{\phi}_n - \phi_0\right) = \frac{n^{-1/2}(R_{10} - \phi R_{01})}{n^{-1}R_{01}} = \frac{n^{-1/2}(W_{10}(\tau) - \phi W_{01}(\tau))}{n^{-1}R_{01}},$$

from which it directly follows using Lemma 3.1 that

$$\sqrt{n}\left(\hat{\phi}_n - \phi_0\right) \to_d \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \frac{\int_0^\tau \left(\phi_0^2 h_{011}(t) + \phi_0 h_{100}(t)\right)\lambda_0(t)dt}{\left(\int_0^\tau h_{01}(t)\lambda_0(t)dt\right)^2}, \tag{49}$$

in agreement with the conclusion of Theorem 3.2, and formulas (42) and (33) in this special case.

Moreover, with the canonical choice $a(u,v) = (u+v)^{-1}$, we have

$$\begin{aligned} H_{011}(t) + H_{100}(t) &= \sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{r}}^2(t) A_{\mathbf{r}}^0(t) A_{\mathbf{r}}^1(t)[A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)] \\ &= \sum_{\mathbf{r}\subset\mathcal{R}} \frac{A_{\mathbf{r}}^0(t) A_{\mathbf{r}}^1(t)}{A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t)} = H_{01}(t), \end{aligned} \tag{50}$$

so that under the null $\phi_0 = 1$, (49) simplifies to

$$\sigma^2 = \frac{1}{\left(\int_0^\tau h_{01}(t)\lambda_0(t)dt\right)}. \tag{51}$$

Under the full cohort case, it has been long known that the Mantel-Haenszel estimator, using $a(u,v) = (u+v)^{-1}$, has the same asymptotic variance as the Maximum Partial Likelihood Estimator (MPLE) at the null. We close this section by noting that this result extends to sampling schemes in general, that is, that (51) is the null asymptotic variance of the MPLE derived in BGL.

In general, we let

$$S_{\mathbf{r}}^{(0)}(\phi,t) = \sum_{i\in\mathbf{r}} \phi^{Z_i(t)} \pi_t(\mathbf{r}|i) = \sum_{k=0}^{\eta} \phi^{\alpha_k} A_{\mathbf{r}}^k(t) \tag{52}$$

$$S_{\mathbf{r}}^{(1)}(\phi,t) = \sum_{i\in\mathbf{r}} Z_i(t)\phi^{Z_i(t)-1} \pi_t(\mathbf{r}|i) = \sum_{k=1}^{\eta} \alpha_k \phi^{\alpha_k - 1} A_{\mathbf{r}}^k(t)$$

$$\text{and} \quad E_{\mathbf{r}}(\phi,t) = \frac{S_{\mathbf{r}}^{(1)}(\phi,t)}{S_{\mathbf{r}}^{(0)}(\phi,t)}, \tag{53}$$

and recall $\alpha_0 = 0$; we apply the convention that $0/0 = 0$.

In the classical case $\eta = 1$, under the null $\phi_0 = 1$,

$$S_{\mathbf{r}}^{(0)}(1,t) = \sum_{i\in\mathbf{r}} \pi_t(\mathbf{r}|i) = A_{\mathbf{r}}^0(t) + A_{\mathbf{r}}^1(t), \quad \text{and} \quad S_{\mathbf{r}}^{(1)}(1,t) = A_{\mathbf{r}}^1(t).$$

Referring now to (3.4) of BGL (where $\beta = 0$ there corresponds to $\phi = 1$ here), since $Z^2 = Z$ when $Z \in \{0,1\}$, we have

$$S_{\mathbf{r}}^{(2)}(1,t) = A_{\mathbf{r}}^1(t).$$

The integrand against the baseline hazard function in (3.10) of BGL, which yields the inverse variance of the MPLE, simplifies in this case to

$$
\begin{aligned}
& \sum_{r\subset\mathcal{R}} \left(\frac{S_{\mathbf{r}}^{(2)}(1,t)}{S_{\mathbf{r}}^{(0)}(1,t)} - \left(\frac{S_{\mathbf{r}}^{(1)}(1,t)}{S_{\mathbf{r}}^{(0)}(1,t)} \right)^2 \right) S_{\mathbf{r}}^{(0)}(t) \\
= & \sum_{r\subset\mathcal{R}} \left(\frac{A_{\mathbf{r}}^1(t)}{A_{\mathbf{r}}^0(t)+A_{\mathbf{r}}^1(t)} - \left(\frac{A_{\mathbf{r}}^1(t)}{A_{\mathbf{r}}^0(t)+A_{\mathbf{r}}^1(t)} \right)^2 \right) [A_{\mathbf{r}}^0(t)+A_{\mathbf{r}}^1(t)] \\
= & \sum_{r\subset\mathcal{R}} \left(\frac{A_{\mathbf{r}}^1(t)[A_{\mathbf{r}}^0(t)+A_{\mathbf{r}}^1(t)]}{A_{\mathbf{r}}^0(t)+A_{\mathbf{r}}^1(t)} - \frac{A_{\mathbf{r}}^1(t)^2}{A_{\mathbf{r}}^0(t)+A_{\mathbf{r}}^1(t)} \right) \\
= & \sum_{r\subset\mathcal{R}} \frac{A_{\mathbf{r}}^1(t)A_{\mathbf{r}}^0(t)}{A_{\mathbf{r}}^0(t)+A_{\mathbf{r}}^1(t)},
\end{aligned}
$$

in agreement with (50), showing the variances of the MPLE in BGL and of the Mantel-Haenszel estimator, at the null, are equal, for sampling in general.

4 Baseline Hazard Estimation

To study the baseline hazard estimate (16), we recall definitions (52) and (53), and impose the following additional conditions.

Condition 5 *The ratio $n(t)/n$ is uniformly bounded away from zero in probability as $n \to \infty$.*

Condition 6 *There exist functions e and ψ such that for all $t \in [0,\tau]$ as $n \to \infty$,*

$$\sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r}) E_{\mathbf{r}}(\phi_0,t) \to_p e(\phi_0,t), \tag{54}$$

and

$$n \sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r})^2 \{S_{\mathbf{r}}^{(0)}(\phi_0,t)\}^{-1} \to_p \psi(\phi_0,t). \tag{55}$$

Letting $t_1 < t_2 < \cdots$ be the collection of all failure times, and $\tilde{\mathcal{R}}_j$ the sampled risk set at failure time t_j, we rewrite the cumulative baseline hazard

estimate (16) as

$$\hat{\Lambda}_n(t,\hat{\phi}_n) = \sum_{t_j \le t} \frac{1}{\sum_{i\in\tilde{\mathcal{R}}_j} \hat{\phi}_n^{Z_i(t_j)} w_i(t_j,\tilde{\mathcal{R}}_j)},$$

where the weights $w_i(t,\mathbf{r})$ are given in (6).

Theorem 4.1 *Let Conditions 1-6 hold, and with $e(\phi_0,u)$ as in (54) set*

$$B(t,\phi_0) = \int_0^t e(\phi_0,u)\lambda_0(u)du.$$

Then $n^{1/2}(\hat{\phi}_n - \phi_0)$ and the process

$$X_n(\cdot) = n^{1/2}\left(\hat{\Lambda}_n(\cdot,\hat{\phi}_n) - \Lambda(\cdot)\right) + n^{1/2}(\hat{\phi}_n - \phi_0)B(\cdot,\phi_0)$$

are asymptotically independent. The limiting distribution of $X_n(\cdot)$ is, with $\psi(\phi_0,t)$ as in (55), that of a mean-zero Gaussian martingale with variance function

$$\omega^2(t,\phi_0) = \int_0^t \psi(\phi_0,u)\lambda_0(u)du.$$

In particular, the scaled difference between the estimated and true integrated hazard

$$\sqrt{n}\left(\hat{\Lambda}_n(\cdot,\hat{\phi}_n) - \Lambda(\cdot)\right)$$

converges weakly as $n\to\infty$ to a mean zero Gaussian process with covariance function

$$\sigma_\Lambda^2(s,t) = \omega^2(s\wedge t) + B(s,\phi_0)\sigma^2 B(t,\phi_0).$$

The function $\sigma_\Lambda^2(s,t)$ can be estimated uniformly consistently by $\hat{\sigma}_\Lambda^2(s,t)$ where

$$\begin{aligned}
\hat{\sigma}_\Lambda^2(s,t) &= \hat{\omega}^2(s\wedge t;\hat{\phi}_n) + \hat{B}_n(s;\hat{\phi}_n)\hat{\sigma}_n^2\hat{B}_n(t;\hat{\phi}_n),\\
\hat{\omega}^2(t;\phi) &= n\sum_{t_j\le t}\frac{1}{\left\{\sum_{i\in\tilde{\mathcal{R}}_j}\phi^{Z_i(t_j)}w_i(t_j,\tilde{\mathcal{R}}_j)\right\}^2}\\
\hat{B}_n(t,\phi) &= \sum_{t_j\le t}\frac{\sum_{i\in\tilde{\mathcal{R}}_j} Z_i(t_j)\phi^{Z_i(t_j)-1}w_i(t_j,\tilde{\mathcal{R}}_j)}{\left\{\sum_{i\in\tilde{\mathcal{R}}_j}\phi^{Z_i(t_j)}w_i(t_j,\tilde{\mathcal{R}}_j)\right\}^2},
\end{aligned}$$

and $\hat{\sigma}_n^2$ is any consistent estimator of σ^2 of (41), such as (45).

Proof: The form of $\hat{\Lambda}_n$ is the same as in BGL, and noting in particular that Condition 4 in BGL can be satisfied by letting $X_{\mathbf{r}}(t) = \max_{0\leq j\leq\eta}\alpha_j$ and $D(t)$ a constant, we have that $X_n(\cdot)$ is asymptotically equivalent to the local square integrable martingale,

$$Y_n(\cdot) = n^{1/2}\int_0^{\cdot}\sum_{\mathbf{r}\subset\mathcal{R}}\frac{dM_{\mathbf{r}}(u)}{\sum_{i\in\mathbf{r}}\phi_0^{Z_i(u)}w_i(u,\mathbf{r})},$$

and the proof of the claims made of the asymptotic distribution of X_n now follow as there.

Regarding the asymptotic independence, note that for any $j<k$, $\mathbf{r}\subset\mathcal{R}$, and locally bounded predictable processes $H_{\mathbf{r}}$,

$$\begin{aligned}
& < \phi_0^{\alpha_k}W_{jk} - \phi_0^{\alpha_j}W_{kj}, \int_0^{\cdot} H_{\mathbf{r}}dM_{\mathbf{r}} >_t \\
= & < \int_0^{\cdot}\sum_{\mathbf{r}} a_{\mathbf{r}}(\phi_0^{\alpha_k}A_{\mathbf{r}}^k dM_{\mathbf{r}}^j - \phi_0^{\alpha_j}A_{\mathbf{r}}^j dM_{\mathbf{r}}^k), \int_0^{\cdot} H_{\mathbf{r}}\sum_{l=0}^{\eta} dM_{\mathbf{r}}^l >_t \\
= & \int_0^t \sum_{\mathbf{r}} a_{\mathbf{r}}\phi_0^{\alpha_k}A_{\mathbf{r}}^k H_{\mathbf{r}} d < M_{\mathbf{r}}^j, M_{\mathbf{r}}^j >_s - \int_0^t\sum_{\mathbf{r}}\phi_0^{\alpha_j}a_{\mathbf{r}}A_{\mathbf{r}}^j H_{\mathbf{r}} d < M_{\mathbf{r}}^k, M_{\mathbf{r}}^k >_s \\
= & \int_0^t\sum_{\mathbf{r}} a_{\mathbf{r}}\left(\phi_0^{\alpha_k}A_{\mathbf{r}}^k H_{\mathbf{r}}\phi_0^{\alpha_j}A_{\mathbf{r}}^j - \phi_0^{\alpha_j}A_{\mathbf{r}}^j H_{\mathbf{r}}\phi_0^{\alpha_k}A_{\mathbf{r}}^k\right)\lambda_0(s)ds = 0.
\end{aligned}$$

Hence, by the asymptotic joint normality provided by Rebolledo's Theorem II.5.1 in Andersen at al. (1993), functions of the collections $\{\int_0^{\cdot} H_{\mathbf{r}}dM_{\mathbf{r}}\}_{\mathbf{r}}$ and $\{\phi_0^{\alpha_k}W_{jk} - \phi_0^{\alpha_j}W_{kj}\}_{j<k}$, in particular $X_n(\cdot)$ and $n^{-1/2}\mathcal{U}_n(\phi_0)$, are asymptotically independent. But by (47), $n^{-1/2}\mathcal{U}_n(\phi_0)$ and a non-zero constant multiple of $\sqrt{n}(\hat{\phi}_n - \phi_0)$ are asymptotically equivalent.

The claim that $\sigma_\Lambda^2(s,t)$ can be estimated uniformly consistently by $\hat{\sigma}_\Lambda^2(s,t)$ follows as in BGL, based on the fact that $\hat{\omega}^2(t,\phi_0)$ is the optional variation process of the local square integrable martingale $Y_n(\cdot)$, which by Rebolledo's theorem as cited above, converges uniformly in probability to its predictable variation $\omega^2(t,\phi_0)$; the uniform convergence of $\hat{B}_n(\cdot,\hat{\phi}_n)$ to $B(\cdot,\phi_0)$ is as in BGL, Proposition 2. ∎

5 Examples

We apply our results to the designs discussed in Section 1, highlighting the classical case where $\eta = 1, \alpha_0 = 0$, and $\alpha_1 = 1$, with the canonical choice of a

given in (11). Though our asymptotic results hold under the weaker stability conditions of Sections 3 and 4, here assume that the censoring, covariate and strata variables are i.i.d. copies of $Y(t), Z(t)$ and $C(t)$ respectively, left continuous and adapted processes having right hand limits. The strata variable needed for Designs 3 and 4 gives the 'type' of individual among the possible values in a (small) finite set $\mathcal{C}$; the strata variable may be used to model any additional information, a surrogate of exposure in particular.

For each of the Designs 1 through 4, we verify that Conditions 1 through 6 are satisfied, and determine the standardized asymptotic distributions of $\hat{\beta}_n$ and $\hat{\Lambda}_n$. We assume that $\tau < \infty$, and so, since λ_0 is already assumed bounded away from infinity, the finite interval Condition 1 holds. As already noted, due to our choice of the (standard) function a as in (11), only the convergence of $h_{n,\mathbf{v}}(t)$ to $h_{\mathbf{v}}(t)$ for $|\mathbf{v}| \in \{2,3\}$ is required in order to satisfy Condition 2. To satisfy Condition 3, letting

$$f_k(t) = P(Z(t) = \alpha_k | Y(t) = 1) \quad \text{for } k = 0, \ldots, \eta,$$

for Designs 1 and 2 we assume that some $j < k$ with $c_{jk} > 0$ there is a non-trivial interval of time $[a,b] \subset [0,\tau]$ over which both $f_j(t)$ and $f_k(t)$ are bounded away from 0. In typical cases, one would have $c_{jk} > 0$ for all pairs $j < k$ in order to take maximum advantage of the available information, and there would be a positive probability in some intervals of time that an at risk individual has covariate α_k; in such a situation any pair $j < k$ can be used to demonstrate the satisfaction of Condition 3.

Let

$$q_l(t) = P(C(t) = l | Y(t) = 1),$$

and

$$f_{k,l}(t) = P(Z(t) = \alpha_k | C(t) = l, Y(t) = 1) \quad k = 0, \ldots, \eta, \ l \in \mathcal{C}.$$

For Design 3, to satisfy Condition 3 we assume that there exists a pair $j < k$ with $c_{jk} > 0$ and $l \in \mathcal{C}$ with $m_l \geq 2$ such that over some non-trivial interval $[a,b] \subset [0,\tau]$ the functions $q_l(t), f_{j,l}(t)$ and $f_{k,l}(t)$ are bounded away from zero. That is, that there is some strata in which a comparison of individuals can be made, and in that strata, the covariate value is not a constant.

For Design 4, to satisfy Condition 3 we assume either i) the assumption for Design 3 holds, or ii) that there exists a pair $j < k$ with $c_{jk} > 0$ and for some unequal pair l_1, l_2 the functions $q_{l_1}(t), q_{l_2}(t), f_{j,l_1}(t), f_{k,l_2}(t)$ are bounded away

from zero. That is, we need to assume either that a meaningful comparison can be drawn i) within a strata or ii) between two different strata. Design 2 is a special case of Designs 3 and 4 with $\mathcal{C} = \{l\}, m_l \geq 2$, and $q_l(t) = 1$ and so i) recovers the assumption in Design 2 used to ensure Condition 3.

As noted above, Condition 4 holds due to our choice of function a. Condition 5 is satisfied using that $\tau < \infty$, and assuming that

$$\inf_{t \in [0,\tau]} p(t) > 0, \quad \text{where} \quad p(t) = P(Y(t) = 1);$$

one needs only to invoke the strong law of large numbers in $D[0,1]$ of Rao (1963) (after reversing the time axis), similar to BGL. We show Condition 6 is satisfied in each of our examples below by proving convergence to, and identifying, the indicated limiting functions. In summary, in each of the examples which follow, we need only verify Conditions 2, 3, and 6. Throughout we let

$$n(t) = |\mathcal{R}(t)| \quad \text{and} \quad \rho_n(t) = n(t)/n.$$

Design 1 *Full Cohort. In this situation all individuals who are at risk at the time of failure are sampled, giving $\pi_t(\mathbf{r}|i) = \mathbf{1}(\mathbf{r} = \mathcal{R}(t))$. Recalling that $n_k(t) = |\mathcal{R}_k(t)|$, the number of individuals in $\mathcal{R}(t)$ with covariate k at time t, we have*

$$A_{\mathbf{r}}^k(t) = \sum_{i \in \mathcal{R}_k(t)} \pi_t(\mathbf{r}|i) = n_k(t)\mathbf{1}(\mathbf{r} = \mathcal{R}(t)).$$

By (12) $a_{\mathcal{R}(t)}(t) = n(t)^{-1}$, and with $\mathcal{T}_j$ the collection of failure times of individuals having exposure j,

$$R_{jk}(\tau) = \int_0^\tau \sum_{\mathbf{r} \subset \mathcal{R}} a_{\mathbf{r}}(t) A_{\mathbf{r}}^k(t) dN_{\mathbf{r}}^j(t) = \int_0^\tau \frac{n_k(t)}{n(t)} dN_{\mathcal{R}(t)}^j(t) = \sum_{t \in \mathcal{T}_j} \frac{n_k(t)}{n(t)},$$

in agreement with (1). Using (27), for $|\mathbf{v}| = 2, 3$,

$$h_{\mathbf{v},n}(t) \;\; = \;\; \frac{1}{n} \sum_{\mathbf{r} \subset \mathcal{R}} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t) \prod_{k \in \mathbf{v}} A_{\mathbf{r}}^k(t) = \rho_n(t) \prod_{k \in \mathbf{v}} \frac{n_k(t)}{n(t)} \rightarrow_p p(t) \prod_{k \in \mathbf{v}} f_k(t) = h_{\mathbf{v}}(t);$$

hence Condition 2 is satisfied. Using that λ_0 is bounded away from zero, Condition 3 is satisfied by the pair $j < k$ for which $f_j(t)$ and $f_k(t)$ are assumed bounded away from zero over some interval.

It remains only to verify Condition 6. By (52) and (53),

$$S^{(0)}_{\mathcal{R}(t)}(\phi,t) = \sum_{k=0}^{\eta} \phi^{\alpha_k} n_k(t) \quad \textit{and} \quad S^{(1)}_{\mathcal{R}(t)}(\phi,t) = \sum_{k=1}^{\eta} \alpha_k \phi^{\alpha_k - 1} n_k(t),$$

so

$$E_{\mathcal{R}(t)}(\phi_0,t) \;=\; \frac{\sum_{k=1}^{\eta} \alpha_k \phi^{\alpha_k-1} n_k(t)}{\sum_{k=0}^{\eta} \phi^{\alpha_k} n_k(t)}.$$

Hence we identify the limiting functions as

$$\sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r}) E_{\mathbf{r}}(\phi_0,t) = \frac{\sum_{k=1}^{\eta} \alpha_k \phi^{\alpha_k-1} n_k(t)}{\sum_{k=0}^{\eta} \phi^{\alpha_k} n_k(t)} \to_p \frac{\sum_{k=1}^{\eta} \alpha_k \phi^{\alpha_k-1} f_k(t)}{\sum_{k=0}^{\eta} \phi^{\alpha_k} f_k(t)} = e(\phi_0,t) \quad \textit{and}$$

$$n \sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r})^2 \{S^{(0)}_{\mathbf{r}}(\phi_0,t)\}^{-1} = \frac{n}{\sum_{k=0}^{\eta} \phi^{\alpha_k} n_k(t)} \to_p \frac{1}{\sum_{k=0}^{\eta} \phi^{\alpha_k} f_k(t)} = \psi(\phi_0,t),$$

thereby fulfilling Condition 6.

In the classical case, we may write the numerator of (49) as

$$\int_0^\tau \left(\phi_0^2 f_1(t) + \phi_0 f_0(t)\right) p(t) f_0(t) f_1(t) \lambda_0(t) dt = \int_0^\tau \left(\phi_0^2 f_1(t) + \phi_0 f_0(t)\right) h_{01}(t) \lambda_0(t) dt$$

and so

$$\sigma^2 = \frac{\int_0^\tau \left(\phi_0^2 f_1(t) + \phi_0 f_0(t)\right) h_{01}(t)\lambda_0(t) dt}{\left(\int_0^\tau h_{01}(t)\lambda_0(t)dt\right)^2}.$$

For the parameters in the asymptotic distribution of the estimate of the baseline hazard, we have in this case

$$e(\phi_0,t) = \frac{f_1(t)}{f_0(t) + \phi_0 f_1(t)} \quad \textit{and} \quad \psi(\phi_0,t) = \frac{1}{f_0(t)+\phi_0 f_1(t)}.$$

Specializing further to the null case $\phi_0 = 1$, we have $f_0(t) + \phi_0 f_1(t) = f_0(t) + f_1(t) = 1$, and hence

$$\sigma^2 = \frac{1}{\int_0^\tau p(t) f_0(t) f_1(t) \lambda_0(t) dt} \tag{56}$$

and

$$e(\phi_0,t) = f_1(t) \quad \textit{and} \quad \psi(\phi_0,t) = 1.$$

In the next three examples we require certain limits of the multivariate hypergeometric distribution

$$\mathbf{X} \sim \mathcal{H}_{\eta+1}(\mathbf{n}, m)$$

having integer parameters $\eta \geq 0, m \geq 0$, and $\mathbf{n} = (n_0, \ldots, n_\eta)$ a vector of non-negative integers, whose j^{th} component X_j counts the number of items of type j contained in a sample without replacement of size m taken from a set having n_j items of type j. That is, for $\mathbf{x} = (x_0, \ldots, x_\eta)$ with non-negative integer components and $|\mathbf{x}| = x_0 + \ldots + x_\eta$,

$$P(\mathbf{X} = \mathbf{x}) = \frac{\prod_{j=0}^{\eta} \binom{n_j}{x_j}}{\binom{n}{m}}, \quad \text{for } |\mathbf{x}| = m \text{ and } |\mathbf{n}| = n. \tag{57}$$

Proposition 5.1 *Let X have distribution (57). If $n_j/n \to f_j \in [0,1]$ for all $j = 0, \ldots, \eta$ as $n \to \infty$, then for all bounded continuous functions G,*

$$EG(\mathbf{X}) \to EG(\mathbf{Y}) \quad \textit{as } n \to \infty, \tag{58}$$

where the vector $\mathbf{Y} \sim M(\mathbf{f}, m)$ has the multinomial distribution

$$P(\mathbf{Y} = \mathbf{x}) = \binom{m}{\mathbf{x}} \mathbf{f}^{\mathbf{x}} \quad \textit{for } |\mathbf{x}| = m \textit{ and } \quad \mathbf{f}^{\mathbf{x}} = \prod_{j=0}^{\eta} f_j^{x_j},$$

whose j^{th} component Y_j counts the number of items of type j included when m items are sampled with replacement from a population where the fraction of type j items is f_j.

In particular, we have convergence of the moments

$$EX_j \to m f_j,$$

$$EX_j X_k \to (m)_2 f_j f_k, \quad EX_j^2 \to m f_j + (m)_2 f_j^2,$$

$$EX_j X_k X_q \to (m)_3 f_j f_k f_q, \quad EX_j^2 X_k \to (m)_2 f_j f_k + (m)_3 f_j^2 f_k,$$

and

$$EX_j^3 \to m f_j + 3(m)_2 f_j^2 + (m)_3 f_j^3.$$

If $\mathbf{n}/n \to_p \mathbf{f}$, a (possibly random) vector of limiting frequencies, then

$$E[G(\mathbf{X})|\mathbf{n}] \to_p E[G(\mathbf{Y})|\mathbf{n}] \quad \textit{as } n \to \infty, \tag{59}$$

and similarly for the convergence of the indicated moments.

Proof: The convergence in distribution, giving (58), of the hypergeometric to the multinomial is well known, and may be shown, for example, by coupling the two distributions so they are equal except on the set of vanishingly small probability where the sample with replacement includes some individual more than once. The convergence of the indicated moments of the hypergeometric to the corresponding moments of the multinomial now follows from the boundedness given by $|\mathbf{X}| = m$.

When $\mathbf{n}/n \to_p \mathbf{f}$, for every subsequence of n there exists a further subsequence where $\mathbf{n}/n \to \mathbf{f}$ almost surely, and the first part of the Lemma gives almost sure convergence of $E[G(\mathbf{X})|\mathbf{n}]$ to $E[G(\mathbf{Y})|\mathbf{n}]$ along this subsequence. Hence the full sequence converges in probability. ■

In what follows we suppress the conditioning in (59) on $\mathbf{n}$.

Design 2 *Simple Random Sampling. The sampling probabilities for this design are given by*

$$\pi_t(\mathbf{r}|i) = \binom{n(t)-1}{m-1}^{-1} \mathbf{1}(\mathbf{r} \ni i, |\mathbf{r}| = m, \mathbf{r} \subset \mathcal{R}(t)),$$

yielding that for $\mathbf{r} \subset \mathcal{R}(t)$ *with* $|\mathbf{r}| = m$ *we have* $\pi_t(\mathbf{r}) = \binom{n(t)}{m}^{-1}$*, and letting* $\mathbf{r}_k(t) = \{i \in \mathbf{r}, Z_i(t) = \alpha_k\}$ *and* $r_k(t) = |\mathbf{r}_k(t)|$,

$$A_{\mathbf{r}}^k(t) = \sum_{i \in \mathcal{R}_k(t)} \pi_t(\mathbf{r}|i) = r_k(t) \binom{n(t)-1}{m-1}^{-1},$$

and by (12)

$$a_{\mathbf{r}}(t) = \frac{1}{m} \binom{n(t)-1}{m-1}.$$

Hence

$$\begin{aligned} h_{n,\mathbf{v}}(t) &= \frac{1}{n} \sum_{\mathbf{r} \subset \mathcal{R}} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t) \prod_{k \in \mathbf{v}} A_{\mathbf{r}}^k(t) = \frac{1}{nm^{|\mathbf{v}|-1}} \binom{n(t)-1}{m-1}^{-1} \sum_{|\mathbf{r}|=m, \mathbf{r} \subset \mathcal{R}(t)} \prod_{k \in \mathbf{v}} r_k(t) \\ &= \frac{n(t)}{nm^{|\mathbf{v}|}} \binom{n(t)}{m}^{-1} \sum_{|\mathbf{r}|=m, \mathbf{r} \subset \mathcal{R}(t)} \prod_{k \in \mathbf{v}} r_k(t) = \frac{\rho_n(t)}{m^{|\mathbf{v}|}} E \prod_{k \in \mathbf{v}} X_k(t), \end{aligned}$$

where $X_k(t)$ *is the* k^{th} *component of the multivariate hypergeometric vector*

$$\mathbf{X}(t) \sim \mathcal{H}_{\eta+1}(\mathbf{n}(t), m) \quad \text{with} \quad \mathbf{n}(t) = (n_0(t), \ldots, n_\eta(t)).$$

Taking limits for j, k distinct, using Proposition 5.1, for $|\mathbf{v}| = 2$

$$h_{jk}(t) = \frac{p(t)}{m^2}(m)_2 f_j(t) f_k(t) = \left(\frac{m-1}{m}\right) p(t) f_j(t) f_k(t),$$

while for $|\mathbf{v}| = 3$,

$$h_{jjk}(t) = \frac{p(t)}{m^3}\left((m)_2 f_j(t) f_k(t) + (m)_3 f_j^2(t) f_k(t)\right),$$

and with j, k, q distinct,

$$h_{jkq}(t) = \frac{p(t)}{m^3}(m)_3 f_j(t) f_k(t) f_q(t);$$

hence Condition 2 is satisfied. Condition 3 is verified here as it was for Design 1.

We begin the verification of Condition 6 by determining the limiting value $e(\phi_0, t)$ of (54). Using (52) and (53),

$$\begin{aligned}\sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r}) E_{\mathbf{r}}(\phi_0, t) &= \sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r}) \frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} A_{\mathbf{r}}^k(t)}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} A_{\mathbf{r}}^k(t)} \\ &= \binom{n(t)}{m}^{-1} \sum_{\mathbf{r}\subset\mathcal{R}(t), |\mathbf{r}|=m} \frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} r_k(t)}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} r_k(t)}.\end{aligned}$$

Writing this expression as an expectation with respect to the multivariate hypergeometric distribution and taking the limit using Proposition 5.1 gives

$$E\left(\frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} X_k(t)}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} X_k(t)}\right) \to_p \sum_{|\mathbf{x}|=m} \left(\frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} x_k}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} x_k}\right) \binom{m}{\mathbf{x}} \mathbf{f}^{\mathbf{x}}(t) = e(\phi_0, t).$$

Similarly, $\psi(\phi_0, t)$ of (55) is the limit

$$\begin{aligned}n \sum_{\mathbf{r}\subset\mathcal{R}} \pi_t^2(\mathbf{r}) \{S_0(\phi_0, t)\}^{-1} &= \frac{m}{\rho_n(t)} E\left(\frac{1}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} X_k(t)}\right) \\ &\to_p \frac{m}{p(t)} \sum_{|\mathbf{x}|=m} \left(\frac{1}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} x_k}\right) \binom{m}{\mathbf{x}} \mathbf{f}^{\mathbf{x}}(t).\end{aligned}$$

Hence, Condition 6 is satisfied.

In the classical case (49) yields

$$\sigma^2 = \frac{\phi_0 \int_0^\tau p(t) f_0(t) f_1(t) \left[(1+\phi_0) + (f_0(t) + \phi_0 f_1(t))(m-2)\right] \lambda_0(t) dt}{(m-1)\left(\int_0^\tau p(t) f_0(t) f_1(t) \lambda_0(t) dt\right)^2}, \quad (60)$$

and the formulas above specialize to

$$\begin{aligned} e(\phi_0, t) &= \sum_{x_0+x_1=m} \frac{x_1}{x_0+\phi_0 x_1} \binom{m}{x_0, x_1} f_0^{x_0}(t) f_1^{x_1}(t), \quad \textit{and} \\ \psi(\phi_0, t) &= \frac{m}{p(t)} \sum_{x_0+x_1=m} \frac{1}{x_0+\phi_0 x_1} \binom{m}{x_0, x_1} f_0^{x_0}(t) f_1^{x_1}(t). \end{aligned}$$

Under the null $\phi_0 = 1$, in the numerator of (60) we have

$$(1+\phi_0) + (f_0(t) + \phi_0 f_1(t))(m-2) \;=\; 2 + (f_0(t) + f_1(t))(m-2) = m,$$

and therefore

$$\sigma^2 = \left(\frac{m}{m-1}\right) \frac{1}{\int_0^\tau p(t) f_0(t) f_1(t) \lambda_0(t) dt},$$

giving an asymptotic relative efficiency of $(m-1)/m$ with respect to the full cohort variance (56), the same relative efficiency as the MPLE, as was expected by the argument supplied at the end of Section 3. Lastly, in the null case

$$e(\phi_0, t) = f_1(t) \quad \textit{and} \quad \psi(\phi_0, t) = p(t)^{-1}.$$

Previous efficiency work used a recursive representation of the factorial moments of the extended hypergeometric distribution Harkness (1965) to derive an asymptotic variance expression for "small strata" case-control data Breslow (1981), Hauck and Donner (1988). The expressions so derived are the special case of (60) when there is a single case per set, a simplification that has not been previously described. Figure 1 shows efficiency curves relative to the maximum partial likelihood estimator (MPLE) as a function of $\log \phi$ *by m when $f_1(t) \equiv .2$. As noted previously in Breslow (1981), the Mantel-Haenszel estimator has high efficiency relative to the MPLE over a fairly large region around the null.*

In the next two designs we consider, for $\mathbf{r} \subset \mathcal{R}(t)$ let

$$\mathbf{r}_{k,l}(t) = \{i \in \mathbf{r} : Z_i(t) = k, C_i(t) = l\}, \qquad r_{k,l}(t) = |\mathbf{r}_{k,l}(t)|,$$

and

$$n_{k,l}(t) = |\mathcal{R}_k(t) \cap \mathcal{C}_l(t)|,$$

the number of individuals having covariate k and type l at time t. With $\mathbf{n}_l(t) = (n_{0,l}(t), \ldots, n_{\eta,l}(t))$, let

$$\mathbf{X}_l(t) \sim H_{\eta+1}(\mathbf{n}_l(t), m_l), \quad l \in \mathcal{C} \tag{61}$$

be independent multivariate hypergeometric vectors.

Design 3 *The sampling probabilities for the matching design are given by*

$$\pi_t(\mathbf{r}|i) = \binom{c_{C_i(t)}(t) - 1}{m_{C_i(t)} - 1}^{-1} \mathbf{1}(\mathbf{r} \subset \mathcal{C}_{C_i(t)}(t), \mathbf{r} \ni i, |\mathbf{r}| = m_{C_i(t)}),$$

where $\mathcal{C}$ is a set of types, $\mathcal{C}_l(t)$ are all those of type $l \in \mathcal{C}$ at risk at time t, of which there are $c_l(t)$.

For $\mathbf{r} \subset \mathcal{C}_l(t)$ with $|\mathbf{r}| = m_l$, we have

$$A_{\mathbf{r}}^k(t) = \sum_{i \in \mathcal{R}_k(t)} \pi_t(\mathbf{r}|i) = \sum_{i \in \mathbf{r}_{k,l}(t)} \binom{c_l(t)}{m_l}^{-1} \frac{c_l(t)}{m_l} = \binom{c_l(t)}{m_l}^{-1} \frac{c_l(t)}{m_l} r_{k,l}(t).$$

Since for such $\mathbf{r}$ we have $\sum_{k=0}^{\eta} r_{k,l}(t) = m_l$, summing over k yields

$$a_{\mathbf{r}}(t) = \frac{1}{c_l(t)} \binom{c_l(t)}{m_l} \mathbf{1}(\mathbf{r} \subset \mathcal{C}_l(t), |\mathbf{r}| = m_l).$$

Hence,

$$\begin{aligned} h_{n,\mathbf{v}}(t) &= \frac{1}{n} \sum_{\mathbf{r} \subset \mathcal{R}} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t) \prod_{k \in \mathbf{v}} A_{\mathbf{r}}^k(t) \\ &= \frac{1}{n} \sum_{l \in \mathcal{C}} \sum_{\mathbf{r} \subset \mathcal{C}_l(t), |\mathbf{r}| = m_l} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t) \prod_{k \in \mathbf{v}} A_{\mathbf{r}}^k(t) \\ &= \frac{n(t)}{n} \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)} \binom{c_l(t)}{m_l}^{-1} \sum_{\mathbf{r} \subset \mathcal{C}_l(t), |\mathbf{r}| = m_l} \prod_{k \in \mathbf{v}} m_l^{-1} r_{k,l}(t). \end{aligned}$$

Then with $\mathbf{X}_l(t)$ as in (61) we can write

$$h_{n,\mathbf{v}}(t) = \rho_n(t) \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)} E \prod_{k \in \mathbf{v}} m_l^{-1} X_{k,l}(t).$$

For $\mathbf{v} = \{k_1, k_2\}$ *distinct, taking limits using Proposition 5.1 we find*

$$h_{\mathbf{v}}(t) = p(t) \sum_{l \in \mathcal{C}} \left(\frac{m_l - 1}{m_l} \right) q_l(t) f_{k_1,l}(t) f_{k_2,l}(t). \quad (62)$$

For $\mathbf{v} = \{k_1, k_1, k_2\}$ *with* k_1, k_2 *distinct, we have*

$$h_{\mathbf{v}}(t) = p(t) \sum_{l \in \mathcal{C}} q_l(t) \left(\frac{m_l - 1}{m_l^2} f_{k_1,l}(t) f_{k_2,l}(t) + \frac{(m_l - 1)_2}{m_l^2} f_{k_1,l}^2(t) f_{k_2,l}(t) \right),$$

and for $\mathbf{v} = \{k_1, k_2, k_3\}$ *all distinct,*

$$h_{\mathbf{v}}(t) = p(t) \sum_{l \in \mathcal{C}} \frac{(m_l - 1)_2}{m_l^2} q_l(t) f_{k_1,l}(t) f_{k_2,l}(t) f_{k_3,l}(t).$$

Hence Condition 2 is satisfied. Condition 3 is satisfied in a manner similarly as for Design 2, with the additional assumption that $m_l \geq 2$*, ensuring that* $(m_l - 1)/m_l$ *in (62) is positive.*

For the verification of Condition 6, we have

$$\pi_t(\mathbf{r}) = \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)} \binom{c_l(t)}{m_l}^{-1} I(\mathbf{r} \subset \mathcal{C}_l(t), |\mathbf{r}| = m_l),$$

so for the limiting value $e(\phi_0, t)$ *of (54) we have*

$$\begin{aligned}
\sum_{\mathbf{r} \subset \mathcal{R}} \pi_t(\mathbf{r}) E_{\mathbf{r}}(\phi_0, t) &= \sum_{\mathbf{r} \subset \mathcal{R}} \pi_t(\mathbf{r}) \frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} A_{\mathbf{r}}^k(t)}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} A_{\mathbf{r}}^k(t)} \\
&= \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)} \binom{c_l(t)}{m_l}^{-1} \sum_{\mathbf{r} \subset \mathcal{C}_l(t), |\mathbf{r}| = m_l} \frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} r_{k,l}(t)}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} r_{k,l}(t)} \\
&= \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)} E \left(\frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} X_{k,l}(t)}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} X_{k,l}(t)} \right) \\
&\to_p \sum_{l \in \mathcal{C}} q_l(t) \sum_{|\mathbf{x}_l| = m_l} \left(\frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} x_{k,l}}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} x_{k,l}} \right) \binom{m_l}{\mathbf{x}_l} \mathbf{f}_l^{\mathbf{x}}(t) = e(\phi_0, t).
\end{aligned}$$

Similarly, $\psi(\phi_0, t)$ *of (55) is the limit*

$$\begin{aligned}
n \sum_{\mathbf{r} \subset \mathcal{R}} \pi_t^2(\mathbf{r}) \{S_{\mathbf{r}}^{(0)}(\phi_0, t)\}^{-1} &= n \sum_{l \in \mathcal{C}} \frac{c_l(t)}{n(t)^2} m_l E \left(\frac{1}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} X_{k,l}(t)} \right) \\
&\to_p p(t)^{-1} \sum_{l \in \mathcal{C}} q_l(t) m_l \sum_{|\mathbf{x}_l| = m_l} \left(\frac{1}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} x_{k,l}} \right) \binom{m_l}{\mathbf{x}_l} \mathbf{f}_l^{\mathbf{x}}(t),
\end{aligned}$$

and Condition 6 is satisfied.

For the classical case, from (49),

$$\sigma^2 = \frac{\phi_0 \int_0^\tau p(t) \sum_{l\in\mathcal{C}} \left(\frac{m_l-1}{m_l^2}\right) q_l(t) f_{0,l}(t) f_{1,l}(t)[(1+\phi_0) + (f_{0,l}(t) + \phi_0 f_{1,l}(t))(m_l-2)]\lambda_0(t)dt}{\left(\int_0^\tau p(t) \sum_{l\in\mathcal{C}} \left(\frac{m_l-1}{m_l}\right) q_l(t) f_{0,l}(t) f_{1,l}(t)\lambda_0(t)dt\right)^2},$$

and the formulas above specialize to

$$e(\phi_0, t) = \sum_{l\in\mathcal{C}} q_l(t) \sum_{x_{0,l}+x_{1,l}=m_l} \left(\frac{x_{1,l}}{x_{0,l}+\phi_0 x_{1,l}}\right) \binom{m_l}{x_{0,l}, x_{1,l}} f_{0,l}^{x_{0,l}}(t) f_{1,l}^{x_{1,l}}(t),$$

and

$$\psi(\phi_0, t) = p(t)^{-1} \sum_{l\in\mathcal{C}} q_l(t) m_l \sum_{x_{0,l}+x_{1,l}=m_l} \left(\frac{1}{x_{0,l}+\phi_0 x_{1,l}}\right) \binom{m_l}{x_{0,l}, x_{1,l}} f_{0,l}^{x_{0,l}}(t) f_{1,l}^{x_{1,l}}(t).$$

Specializing further, under the null $\phi_0 = 1$,

$$\sigma^2 = \frac{1}{\left(\int_0^\tau p(t) \sum_{l\in\mathcal{C}} \left(\frac{m_l-1}{m_l}\right) q_l(t) f_{0,l}(t) f_{1,l}(t)\lambda_0(t)dt\right)},$$

$$e(\phi_0, t) = \sum_{l\in\mathcal{C}} q_l(t) f_{1,l}(t), \quad \text{and} \quad \psi(\phi_0, t) = p(t)^{-1}.$$

Remaining in the classical case, we more generally consider the matching framework where each strata $l \in \mathcal{C}$ has its own baseline $\lambda_l(t)$, so that if $C_i(t) \in \mathcal{C}$ is the strata of individual i at time t, the observed failure intensity for individual i is given by

$$\lambda_i(t) = Y_i(t)\phi_0^{Z_i(t)}\lambda_{C_i(t)}(t).$$

Even in this extended model, it remains true that

$$R_{10}(t) - \phi_0 R_{01}(t)$$

is a local square integrable martingale. We can guarantee the consistency of $\hat{\phi}_n$ by letting Condition 1 hold with $\lambda_l(t)$ replacing $\lambda_0(t)$, and Condition 2 hold with

$$H_{\mathbf{v},l}(t) = \sum_{r\subset\mathcal{C}_l(t)} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t) \prod_{k\in\mathbf{v}} A_{\mathbf{r}}^k(t)$$

and its scaled limit $h_{\mathbf{v},l}(t)$ replacing and $H_{\mathbf{v}}(t)$ and $h_{\mathbf{v}}(t)$ respectively, for each $l \in \mathcal{C}$. In addition, when Condition 4 holds for each $\mathcal{C}_l(t)$ replacing $\mathcal{R}$ for each $l \in \mathcal{C}$, the asymptotic variance of $\hat{\phi}_n$ for the matching design with strata specific baseline hazard $\lambda_l(t)$ is given by

$$\sigma^2 = \frac{\int_0^\tau \sum_{l\in\mathcal{C}} (\phi_0^2 h_{011,l}(t) + \phi_0 h_{100,l}(t))\lambda_l(t)dt}{(\int_0^\tau \sum_{l\in\mathcal{C}} h_{01,l}(t)\lambda_l(t)dt)^2}.$$

Design 4 *The sampling probabilities for the counter matching design are given by*

$$\pi_t(\mathbf{r}|i) = \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1} \frac{c_{C_i(t)}(t)}{m_{C_i(t)}} \mathbf{1}(\mathbf{r} \ni i, \mathbf{r} \in \mathcal{P}_{\mathcal{C}}(t)),$$

where $\mathcal{C}$ is a set of types, $\mathcal{P}_{\mathcal{C}}(t) \subset \mathcal{R}(t)$ the collection of sets $\mathbf{r}$ with m_l subjects of type l at time t, $c_l(t)$ is the number of type l subjects in $\mathcal{R}(t)$, and $C_i(t)$ the type of subject i at time t; by (4),

$$\pi_t(\mathbf{r}) = \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1} \mathbf{1}(\mathbf{r} \in \mathcal{P}_{\mathcal{C}}(t)). \tag{63}$$

Letting $\mathbf{r}_{k,l}(t) = \{i \in \mathbf{r} : Z_i(t) = k, C_i(t) = l\}$, and $r_{k,l}(t) = |\mathbf{r}_{k,l}(t)|$ we have

$$A_{\mathbf{r}}^k(t) = \sum_{i\in\mathcal{R}_k(t)} \pi_t(\mathbf{r}|i) = \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1} \left(\sum_{l\in\mathcal{C}} r_{k,l}(t)\frac{c_l(t)}{m_l}\right) \mathbf{1}(\mathbf{r} \in \mathcal{P}_{\mathcal{C}}(t)),$$

and for $\mathbf{r} \in \mathcal{P}_{\mathcal{C}}(t)$, by (12),

$$a_{\mathbf{r}}(t) = \frac{1}{n(t)} \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right] \mathbf{1}(\mathbf{r} \in \mathcal{P}_{\mathcal{C}}(t)).$$

Hence,

$$\begin{aligned} h_{n,\mathbf{v}}(t) &= \frac{1}{n}\sum_{\mathbf{r}\subset\mathcal{R}} a_{\mathbf{r}}^{|\mathbf{v}|-1}(t) \prod_{k\in\mathbf{v}} A_{\mathbf{r}}^k(t) \\ &= \frac{n(t)}{n} \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1} \sum_{\mathbf{r}\in\mathcal{P}_{\mathcal{C}}(t)} \prod_{k\in\mathbf{v}} \left(\sum_{l\in\mathcal{C}} \frac{r_{k,l}(t)}{m_l}\frac{c_l(t)}{n(t)}\right). \end{aligned}$$

Now we can write the above sum as the expectation

$$h_{n,\mathbf{v}}(t) = \rho_n(t)E\left(\prod_{k\in\mathbf{v}}\sum_{l\in\mathcal{C}}\frac{X_{k,l}(t)c_l(t)}{m_l n(t)}\right) = \rho_n(t)E\left(\sum_{l_p\in\mathcal{C},p=1,\ldots,|\mathbf{v}|}\prod_{k\in\mathbf{v}}\frac{X_{k,l_p}(t)c_{l_p}(t)}{m_{l_p}n(t)}\right). \quad (64)$$

For the case $|\mathbf{v}| = 2$ with $\mathbf{v} = \{k_1, k_2\}$ distinct, the expectation in (64) expands into the diagonal and off diagonal sums,

$$E\left(\sum_{l\in\mathcal{C}}\frac{X_{k_1,l}(t)X_{k_2,l}(t)c_l^2(t)}{m_l^2 n^2(t)} + \sum_{l_1\neq l_2}\frac{X_{k_1,l_1}(t)X_{k_2,l_2}(t)c_{l_1}(t)c_{l_2}(t)}{m_{l_1}m_{l_2}n^2(t)}\right).$$

Letting

$$\frac{c_l(t)}{n(t)} \to_p q_l(t) \quad \textit{and} \quad \frac{n_{k,l}(t)}{c_l(t)} \to_p f_{k,l}(t),$$

and applying Proposition 5.1 we find that $h_{n,\mathbf{v}}(t)$ converges to $p(t)$ times

$$\begin{aligned}
&\sum_{l\in\mathcal{C}}\frac{(m_l)_2 f_{k_1,l}(t)f_{k_2,l}(t)q_l^2(t)}{m_l^2} + \sum_{l_1\neq l_2} f_{k_1,l_1}(t)f_{k_2,l_2}(t)q_{l_1}(t)q_{l_2}(t) \\
=\ &\sum_{l\in\mathcal{C}}\left(\frac{m_l-1}{m_l}\right) f_{k_1,l}(t)f_{k_2,l}(t)q_l^2(t) + \sum_{l_1\neq l_2} f_{k_1,l_1}(t)f_{k_2,l_2}(t)q_{l_1}(t)q_{l_2}(t) \quad (65) \\
=\ &\left(\sum_{l\in\mathcal{C}} f_{k_1,l}(t)q_l(t)\right)\left(\sum_{l\in\mathcal{C}} f_{k_2,l}(t)q_l(t)\right) - \sum_{l\in\mathcal{C}}\left(\frac{1}{m_l}\right) f_{k_1,l}(t)f_{k_2,l}(t)q_l^2(t) \\
=\ & f_{k_1}(t)f_{k_2}(t) - \sum_{l\in\mathcal{C}}\left(\frac{1}{m_l}\right) f_{k_1,l}(t)f_{k_2,l}(t)q_l^2(t). \quad (66)
\end{aligned}$$

Applying the assumptions made at the beginning of this section in version i) on the first sum in (65), or in version ii) on the second sum in (65), we find Condition 3 satisfied.

For $|\mathbf{v}| = 3$ we consider the two cases $\mathbf{v} = \{k_1, k_1, k_3\}$ with $k_1 \neq k_3$ and $\mathbf{v} = \{k_1, k_2, k_3\}$, all distinct. In the first case, applying Proposition 5.1, for the diagonal term $l_1 = l_2 = l_3$ in expression (64),

$$E\left(\sum_{l\in\mathcal{C}}\frac{X_{k_1,l}^2(t)X_{k_3,l}(t)c_l^3(t)}{m_l^3 n^3(t)}\right) \to_p \sum_{l\in\mathcal{C}}\frac{[(m_l)_2 f_{k_1,l}(t)f_{k_3,l}(t) + (m_l)_3 f_{k_1,l}^2(t)f_{k_3,l}(t)]q_l^3(t)}{m_l^3},$$

for $l_1 = l_2 \neq l_3$,

$$E\left(\sum_{l_1\neq l_3}\frac{X^2_{k_1,l_1}(t)X_{k_3,l_3}(t)c^2_{l_1}(t)c_{l_3}(t)}{m^2_{l_1}m_{l_3}n^3(t)}\right)\to_p \sum_{l_1\neq l_3}\frac{[m_{l_1}f_{k_1,l_1}(t)+(m_{l_1})_2f^2_{k_1,l_1}(t)]f_{k_3,l_3}(t)q^2_{l_1}(t)q_{l_3}(t)}{m^2_{l_1}},$$

for $l_1 = l_3 \neq l_2$,

$$E\left(\sum_{l_1\neq l_2}\frac{X_{k_1,l_1}(t)X_{k_1,l_2}(t)X_{k_3,l_1}(t)c^2_{l_1}(t)c_{l_2}(t)}{m^2_{l_1}m_{l_2}n^3(t)}\right)\to_p \sum_{l_1\neq l_2}\frac{(m_{l_1})_2f_{k_1,l_1}(t)f_{k_3,l_1}(t)f_{k_1,l_2}(t)q^2_{l_1}(t)q_{l_2}(t)}{m^2_{l_1}},$$

for $l_2 = l_3 \neq l_1$,

$$E\left(\sum_{l_1\neq l_2}\frac{X_{k_1,l_1}(t)X_{k_1,l_2}(t)X_{k_3,l_2}(t)c_{l_1}(t)c^2_{l_2}(t)}{m_{l_1}m^2_{l_2}n^3(t)}\right)\to_p \sum_{l_1\neq l_2}\frac{(m_{l_2})_2f_{k_1,l_1}(t)f_{k_1,l_2}(t)f_{k_3,l_2}(t)q_{l_1}(t)q^2_{l_2}(t)}{m^2_{l_2}},$$

and for l_1, l_2, l_3 *distinct,*

$$\begin{aligned} &E\left(\sum_{|\{l_1,l_2,l_3\}|=3}\frac{X_{k_1,l_1}(t)X_{k_1,l_2}(t)X_{k_3,l_3}(t)c_{l_1}(t)c_{l_2}(t)c_{l_3}(t)}{m_{l_1}m_{l_2}m_{l_3}n^3(t)}\right)\\ \to_p &\sum_{|\{l_1,l_2,l_3\}|=3} f_{k_1,l_1}(t)f_{k_1,l_2}(t)f_{k_3,l_3}(t)q_{l_1}(t)q_{l_2}(t)q_{l_3}(t).\end{aligned}$$

Summing and simplifying, we find that for $\mathbf{v} = \{k_1, k_1, k_3\}$ *with* $k_1 \neq k_3$, $h_{\mathbf{v}}(t)$ *is* $p(t)$ *times*

$$\begin{aligned} &f^2_{k_1}(t)f_{k_3}(t) \\ +\ &\sum_{l\in\mathcal{C}}\frac{1}{m^2_l}f_{k_1,l}(t)f_{k_3,l}(t)\left(m_l(1-3f_{k_1,l})-(1-2f_{k_1,l})\right)q^3_l(t)\\ +\ &\sum_{l_1\neq l_2}\left(\frac{1}{m_{l_1}}\right)f_{k_1,l_1}(t)\left[f_{k_3,l_2}(t)(1-f_{k_1,l_1}(t))-2f_{k_3,l_1}(t)f_{k_1,l_2}(t)\right]q^2_{l_1}(t)q_{l_2}(t).\end{aligned} \tag{67}$$

Similarly, for $\mathbf{v} = \{k_1, k_2, k_3\}$ *distinct, applying Proposition 5.1, we have for* $l_1 = l_2 = l_3$,

$$E\left(\sum_{l\in\mathcal{C}}\frac{X_{k_1,l}(t)X_{k_2,l}(t)X_{k_3,l}(t)c^3_l(t)}{m^3_l n^3(t)}\right)\to_p \sum_{l\in\mathcal{C}}\frac{(m_l)_3f_{k_1,l}(t)f_{k_2,l}(t)f_{k_3,l}(t)q^3_l(t)}{m^3_l},$$

for $l_1 = l_2 \neq l_3$,

$$E\left(\sum_{l_1 \neq l_3} \frac{X_{k_1,l_1}(t)X_{k_2,l_1}(t)X_{k_3,l_3}(t)c_{l_1}^2(t)c_{l_3}(t)}{m_{l_1}^2 m_{l_3} n^3(t)}\right) \to_p \sum_{l_1 \neq l_3} \frac{(m_{l_1})_2 f_{k_1,l_1}(t)f_{k_2,l_1}(t)f_{k_3,l_3}(t)q_{l_1}^2(t)q_{l_3}(t)}{m_{l_1}^2}$$

for $l_1 = l_3 \neq l_2$,

$$E\left(\sum_{l_1 \neq l_2} \frac{X_{k_1,l_1}(t)X_{k_2,l_2}(t)X_{k_3,l_1}(t)c_{l_1}^2(t)c_{l_2}(t)}{m_{l_1}^2 m_{l_2} n^3(t)}\right) \to_p \sum_{l_1 \neq l_2} \frac{(m_{l_1})_2 f_{k_1,l_1}(t)f_{k_2,l_2}(t)f_{k_3,l_1}(t)q_{l_1}^2(t)q_{l_2}(t)}{m_{l_1}^2},$$

for $l_2 = l_3 \neq l_1$,

$$E\left(\sum_{l_1 \neq l_2} \frac{X_{k_1,l_1}(t)X_{k_2,l_2}(t)X_{k_3,l_2}(t)c_{l_1}(t)c_{l_2}^2(t)}{m_{l_1} m_{l_2}^2 n^3(t)}\right) \to_p \sum_{l_1 \neq l_2} \frac{(m_{l_2})_2 f_{k_1,l_1}(t)f_{k_2,l_2}(t)f_{k_3,l_2}(t)q_{l_1}(t)q_{l_2}^2(t)}{m_{l_2}^2},$$

and for l_1, l_2, l_3 *distinct,*

$$\begin{aligned} & E\left(\sum_{|\{l_1,l_2,l_3\}|=3} \frac{X_{k_1,l_1}(t)X_{k_2,l_2}(t)X_{k_3,l_3}(t)c_{l_1}(t)c_{l_2}(t)c_{l_3}(t)}{m_{l_1}m_{l_2}m_{l_3}n^3(t)}\right) \\ \to_p & \sum_{|\{l_1,l_2,l_3\}|=3} f_{k_1,l_1}(t)f_{k_2,l_2}(t)f_{k_3,l_3}(t)q_{l_1}(t)q_{l_2}(t)q_{l_3}(t). \end{aligned}$$

Summing, we find that for $\mathbf{v} = \{k_1, k_2, k_3\}$ *distinct,* $h_{\mathbf{v}}(t)$ *is* $p(t)$ *times*

$$\begin{aligned} & f_{k_1}(t)f_{k_2}(t)f_{k_3}(t) \\ + & \sum_{l \in \mathcal{C}} \left(\frac{-3m_l + 2}{m_l^2}\right) f_{k_1,l}(t)f_{k_2,l}(t)f_{k_3,l}(t)q_l^3(t) \\ - & \sum_{l_1 \neq l_3} \left(\frac{1}{m_{l_1}}\right) f_{k_1,l_1}(t)f_{k_2,l_1}(t)f_{k_3,l_3}(t)q_{l_1}^2(t)q_{l_3}(t) \\ - & \sum_{l_1 \neq l_2} \left(\frac{1}{m_{l_1}}\right) f_{k_3,l_1}(t)[f_{k_1,l_1}(t)f_{k_2,l_2}(t) + f_{k_1,l_2}(t)f_{k_2,l_1}(t)]q_{l_1}^2(t)q_{l_2}(t). \end{aligned}$$

Hence the remaining $|\mathbf{v}| = 3$ *portion of Condition 2 is satisfied.*

For the parameters in the limiting distribution of the estimator of the cumulative baseline hazard, using (63) and applying Proposition 5.1 for each

$l \in \mathcal{C}$ yields

$$\begin{aligned}
\sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r}) E_{\mathbf{r}}(\phi_0, t) &= \sum_{\mathbf{r}\subset\mathcal{R}} \pi_t(\mathbf{r}) \frac{\sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k - 1} A_{\mathbf{r}}^k(t)}{\sum_{k=0}^{\eta} \phi_0^{\alpha_k} A_{\mathbf{r}}^k(t)} \\
&= \left[\prod_{l\in\mathcal{C}} \binom{c_l(t)}{m_l}\right]^{-1} \sum_{\mathbf{r}\in\mathcal{P}_{\mathcal{C}}(t)} \frac{\sum_{k=1}^{\eta} \sum_{l\in\mathcal{C}} \alpha_k \phi_0^{\alpha_k-1} r_{k,l}(t) \frac{c_l(t)}{m_l}}{\sum_{k=0}^{\eta} \sum_{l\in\mathcal{C}} \phi_0^{\alpha_k} r_{k,l}(t) \frac{c_l(t)}{m_l}} \\
&= E\left(\frac{\sum_{l\in\mathcal{C}} \sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k-1} X_{k,l}(t) \frac{c_l(t)}{m_l}}{\sum_{l\in\mathcal{C}} \sum_{k=0}^{\eta} \phi_0^{\alpha_k} X_{k,l}(t) \frac{c_l(t)}{m_l}}\right) \\
&\to_p \sum_{|\mathbf{x}_\alpha| = m_\alpha, \alpha\in\mathcal{C}} \left(\frac{\sum_{l\in\mathcal{C}} \sum_{k=1}^{\eta} \alpha_k \phi_0^{\alpha_k-1} x_{k,l} \frac{q_l(t)}{m_l}}{\sum_{l\in\mathcal{C}} \sum_{k=0}^{\eta} \phi_0^{\alpha_k} x_{k,l} \frac{q_l(t)}{m_l}}\right) \prod_{\alpha\in\mathcal{C}} \binom{m_\alpha}{\mathbf{x}_\alpha} \mathbf{f}_\alpha^{\mathbf{x}}(t) = e(\phi_0, t).
\end{aligned}$$

Similarly, $\psi(\phi_0, t)$ of (55) is obtained by taking the limit

$$\begin{aligned}
n \sum_{\mathbf{r}\subset\mathcal{R}} \pi_t^2(\mathbf{r}) \{S_0(\phi_0, t)\}^{-1} &= \frac{1}{\rho_n(t)} E\left(\frac{1}{\sum_{l\in\mathcal{C}} \sum_{k=0}^{\eta} \phi_0^{\alpha_k} X_{k,l}(t) \frac{c_l(t)}{m_l}}\right) \\
&\to_p \frac{1}{p(t)} \sum_{|\mathbf{x}_\alpha| = m_\alpha, \alpha\in\mathcal{C}} \left(\frac{1}{\sum_{l\in\mathcal{C}} \sum_{k=0}^{\eta} \phi_0^{\alpha_k} x_{k,l} \frac{q_l(t)}{m_l}}\right) \prod_{\alpha\in\mathcal{C}} \binom{m_\alpha}{\mathbf{x}_\alpha} \mathbf{f}_\alpha^{\mathbf{x}}(t).
\end{aligned}$$

Hence, Condition 6 is satisfied.

Specializing to the classical case, the functions $h_{01}(t)$ and $h_{011}(t)$ can be determined from (66) and (67) for $k_1 = 1, k_3 = 0$ respectively, and after some simplification using $1 - f_{k_1,l_1}(t) = f_{k_2,l_1}(t)$ to obtain the following slightly more agreeable form for the latter, we have

$$h_{01}(t) = p(t)\left(f_0(t) f_1(t) - \sum_{l\in\mathcal{C}} \left(\frac{1}{m_l}\right) f_{0,l}(t) f_{1,l}(t) q_l^2(t)\right) \quad \text{and} \tag{68}$$

$$\begin{aligned}
h_{011}(t) = p(t) \big(&f_0(t) f_1^2(t) \\
&+ \left(\sum_{l\in\mathcal{C}} (\frac{1}{m_l}) f_{0,l}(t) f_{1,l}(t) q_l^2(t)\right) \left(\sum_{l\in\mathcal{C}} (1 - 3 f_{1,l}(t)) q_l(t)\right) \\
&- \sum_{l\in\mathcal{C}} (\frac{1}{m_l^2}) f_{0,l}(t) f_{1,l}(t) (1 - 2 f_{1,l}(t)) q_l^3(t)\Big);
\end{aligned}$$

$h_{100}(t)$ is the same as $h_{110}(t)$ with the roles of 0 and 1 reversed. The value of σ^2 can now be calculated by (49).

For the parameters in the limiting distribution for the baseline hazard estimator, we have

$$e(\phi_0,t) = \sum_{x_{0,\alpha}+x_{1,\alpha}=m_\alpha,\alpha\in\mathcal{C}} \left(\frac{\sum_{l\in\mathcal{C}} x_{1,l}\frac{q_l(t)}{m_l}}{\sum_{l\in\mathcal{C}}(x_{0,l}+\phi_0 x_{k,l})\frac{q_l(t)}{m_l}}\right) \prod_{\alpha\in\mathcal{C}} \binom{m_\alpha}{x_{0,\alpha},x_{1,\alpha}} f_\alpha^{x_0}(t) f_\alpha^{x_1}(t),$$

and

$$\psi(\phi_0,t) = p(t)^{-1} \sum_{x_{0,\alpha}+x_{1,\alpha}=m_\alpha,\alpha\in\mathcal{C}} \left(\frac{1}{\sum_{l\in\mathcal{C}}(x_{0,l}+\phi_0 x_{k,l})\frac{q_l(t)}{m_l}}\right) \prod_{\alpha\in\mathcal{C}} \binom{m_\alpha}{x_{0,\alpha},x_{1,\alpha}} f_\alpha^{x_0}(t) f_\alpha^{x_1}(t).$$

We specialize further to the case where there are two strata, $|\mathcal{C}| = 2$, *and the binary strata variable* $C(t) \in \{0,1\}$ *is a (perhaps easily available) surrogate for the true binary exposure* $Z(t) \in \{0,1\}$*. Recalling*

$$f_{k,l}(t) = P(Z(t)=k|C(t)=l, Y(t)=1) \quad k,l \in \{0,1\},$$

we have

$$\begin{aligned} f_{k,l}(t)q_l(t) &= P(Z(t)=k|C(t)=l,Y(t)=1)P(C(t)=l|Y(t)=1) \\ &= P(Z(t)=k, C(t)=l|Y(t)=1) = \pi_{k,l}(t) \quad \textit{say}, \end{aligned}$$

and

$$\delta(t) = P(C(t)=1|Z(t)=1,Y(t)=1) \quad \textit{and} \quad \gamma(t) = P(C(t)=0|Z(t)=0,Y(t)=1),$$

the sensitivity and specificity of $Z(t)$ *for* $C(t)$*. Since*

$$\begin{aligned} \pi_{11}(t) &= \delta(t)f_1(t), & \pi_{10}(t) &= (1-\delta(t))f_1(t) \\ \pi_{01}(t) &= (1-\gamma(t))f_0(t), & \pi_{00}(t) &= \gamma(t)f_0(t), \end{aligned}$$

we can write the expression in (68) in parenthesis for, say $m_0 = m_1 = 1$, *as*

$$\begin{aligned} & f_0(t)f_1(t) - (f_{0,1}(t)f_{1,1}(t)q_1^2(t) + f_{0,0}(t)f_{1,0}(t)q_0^2(t)) \\ =\ & f_0(t)f_1(t) - (\pi_{0,1}(t)\pi_{1,1}(t) + \pi_{0,0}(t)\pi_{1,0}(t)) \\ =\ & f_0(t)f_1(t) - ((1-\gamma(t))f_0(t)\delta(t)f_1(t) + \gamma(t)f_0(t)(1-\delta(t))f_1(t)) \\ =\ & f_0(t)f_1(t)((1-\delta(t))(1-\gamma(t)) + \gamma(t)\delta(t)). \end{aligned} \tag{69}$$

In a similar way $h_{011}(t)$ *and* $h_{001}(t)$ *can be expressed in terms of the sensitivity, specificity, and probability of exposure integrated against the baseline hazard. Using (46) and the partial likelihood variance given in (A3)*

from Langholz and Borgan (1995), asymptotic efficiencies for the Mantel-Haenszel relative to the partial likelihood can be computed. Figure 2 shows the asymptotic relative efficiencies by $\log(\phi)$ *with* $P(Z(t)=1|Y(t)=1)=.2$ *for* $m_0, m_1 \in \{1,2\}$ *when the conditional distribution of* $(Z(t), C(t))$ *given* $Y(t)=1$ *does not depend on* t*, which holds, approximately, for rare outcomes when censoring does not depend on* $(Z(t), C(t))$*. Although there is some difference in the relative efficiencies by choice of* m_0 *and* m_1 *and the sensitivity and specificity of* C *for* Z*, the Mantel-Haenszel estimator has fairly high efficiency in a wide range of situations.*

Under the null $\phi_0 = 1$ *in the classical case, the numerator of the variance formula (49) simplifies since*

$$h_{011}(t) + h_{110}(t) = h_{01}(t),$$

yielding

$$\sigma^2 = \frac{1}{\int_0^\tau p(t)\left(f_0(t)f_1(t) - \sum_l (\frac{1}{m_l}) f_{0,l}(t) f_{1,l}(t) q_l^2(t)\right)\lambda_0(t)}. \tag{70}$$

Under the null in general, using that $\sum_{k=0}^{\eta} x_{k,l} = m_l$ *and* $EX_{k,l} = m_l f_{k,l}(t)$*, we have*

$$e(\phi_0, t) = \sum_{k=1}^{\eta} \sum_{l \in \mathcal{C}} \alpha_k f_{k,l}(t) q_l(t) \quad \text{and} \quad \psi(\phi_0, t) = p(t)^{-1},$$

so in the classical case in particular

$$e(\phi_0, t) = \sum_{l \in \mathcal{C}} f_{1,l}(t) q_l(t).$$

When $(m_0, m_1) = (1,1)$*, so that the design matches one control with 'surrogate exposure'* $C(t)$ *value opposite to the exposure* $Z(t)$ *of the case, substituting (69) into (70) yields*

$$\sigma^2 = \left(\int_0^\tau p(t) f_0(t) f_1(t)((1-\delta(t))(1-\gamma(t)) + \gamma(t)\delta(t))\lambda_0(t)dt\right)^{-1},$$

which is the equal to the asymptotic variance for the $(1,1)$ *counter matching design when using the maximum partial likelihood estimator Langholz and Clayton (1994), a result expected based on the argument at the end of Section 3. We note that, as in Langholz and Clayton (1994), when the sensitivity and specificity are close to 1 (or zero), the counter matching design has efficiency close to that of the full cohort.*

References

[1] Aitchison, J. and Silvey, S.D. (1958) Maximum Likelihood Estimation Subject to Restraints *Ann Math Stat.* **29**, pp 813-828

[2] Andersen, P.K., Borgan, Ø., Gill, R., and Keiding, N. (1993) Statistical Models Based on Counting Processes. Springer Series in Statistics.

[3] Billingsely, P. (1961) Statistical Inference for Markov Processes. University of Chicago Press.

[4] Ø. Borgan, L. Goldstein, and B. Langholz. (1995) Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Annals of Statistics*, 23:1749–1778.

[5] N. Breslow. (1981) Odds ratio estimators when the data are sparse. *Biometrika*, 68:73–84.

[6] N.E. Breslow. (1996) Statistics in epidemiology: The case-control study. *Journal of the American Statistical Society*, 91(433):14–28.

[7] Cox, D.R, and Snell, E.J. (1989) *Analysis of Binary Data.* Monographs on Statistics and Applied Probability 32. Chapman and Hall, New York, second edition.

[8] Godambe, V. (1960) An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208-1211.

[9] W.L. Harkness. (1965) Properties of the extended hypergeometric distribution. *Annals of Mathematical Statistics*, 36(3):938–945.

[10] W. W. Hauck and A. Donner. (1988) The asymptotic relative efficiency of the mantel-haenszel estimator in the increasing-number-of-strata case. *Biometrics*, 44:379–384.

[11] Heyde, C. (1997) Quasi-likelihood and its application. A general approach to optimal parameter estimation. Springer Series in Statistics. Springer-Verlag, New York.

[12] Hjort, N. L. and Pollard, D. (1993). Asymptotics of minimisers of convex processes. Statistical Research Report 5/93, Institute of Mathematics, University of Oslo.

[13] Kahn, H, A. and Sempos, C, T. (1989) Statistical Methods in Epidemiology New York, Oxford University Press

[14] B. Langholz and Ø. Borgan. (1995) Counter-matching: A stratified nested case-control sampling method. *Biometrika*, 82:69–79.

[15] B. Langholz and D. Clayton. (1994) Sampling strategies in nested case-control studies. *Environmental Health Perspectives*, 102(Suppl 8):47–51.

[16] B. Langholz and L. Goldstein. (1996) Risk set sampling in epidemiologic cohort studies. *Statistical Science*, 11:35-53.

[17] Liang, K.Y. (1985) Odds ratio inference with dependent data. *Biometrika*, **72**, 678-682

[18] Mantel N. and Haenszel W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease *Journal of the National Cancer Institute*, **22**, 719-748.

[19] Qu, A., Lindsay, B., and Li, B.(2000) Improving generalised estimating equations using quadratic inference functions. *Biometrika*, **87**, 823-836.

[20] Rao, R.R. (1963) The law of large numbers for $D[0,1]$-valued random variables. *Theory Probab. Appl.*, **8**, 70-74.

[21] Rebolledo, R. (1980) Central limit theorems for local martingales. *Z. Wahrsch. Gebiete*, **51**, pp. 269-286.

[22] J.M. Robins, N. Breslow, and S. Greenland. (1986) Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics*, 42:311-323.

[23] Z-Z. Zhang. (2000) On consistency of Mantel-Haenszel type estimators in nested case-control studies. *Journal of the Japan Statistical Society*, 30(2):205–211.

[24] Z-Z. Zhang, Y. Fujii, and T. Yanagawa. (2000) On Mantel-Haenszel type estimators in simple nested case-control studies. *Communications in Statistics, Part A – Theory and Methods [Split from: @J(CommStat)]*, 29(11):2507–2521.

Figure 1: Asymptotic efficiency by exposure rate ratio ϕ of Mantel-Haenszel relative to the partial likelihood estimator for simple random sampling of $m-1$ controls. Probability of exposure $P(Z(t)=1|Y(t)=1)=.2$.

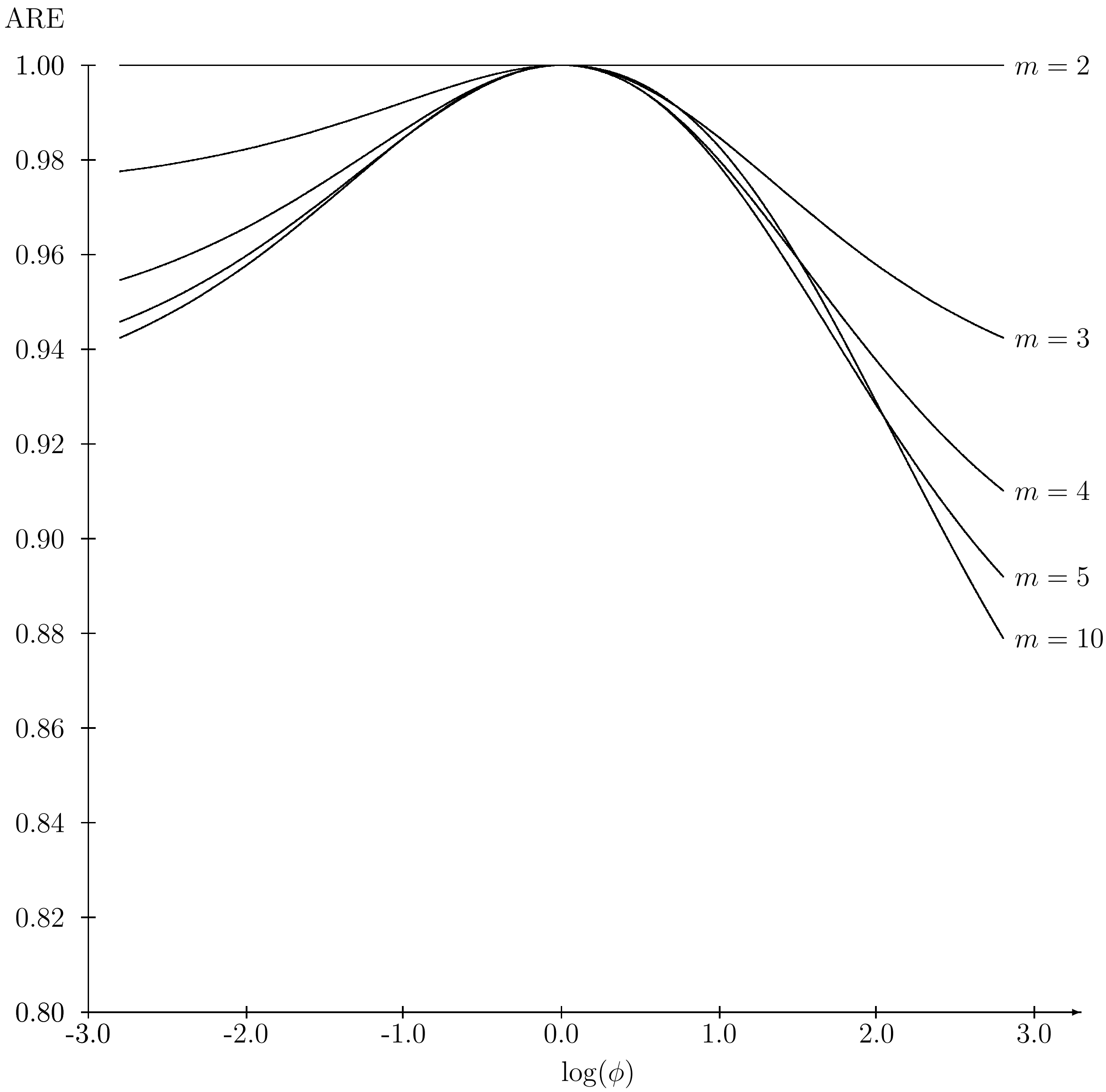

Figure 2: Asymptotic efficiency by exposure rate ratio ϕ of Mantel-Haenszel relative to the partial likelihood estimator for counter-matching by sensitivity ($\delta = P(Z(t) = 1|C(t) = 1, Y(t) = 1)$) and specificity ($\gamma = P(Z(t) = 0|C(t) = 0, Y(t) = 1)$). Probability of exposure $P(Z(t) = 1|Y(t) = 1) = .2$.